

UNIVERSITY  
OF MICHIGAN

PSY

Vol. 49, No. 3

JUN 13 1952

May, 1952

SCIENCE  
LIBRARY

# Psychological Bulletin

---

## CONTENTS

### ARTICLES:

*The Nature and Measurement of Meaning:* CHARLES E. OSGOOD, 197

*The Validity of the Blacky Pictures:* GERALD S. BAUM AND HOWARD F. HUNT, 238

*A Measure of Relation Determined by Both Mean Difference and Profile Information:* CHARLES E. OSGOOD AND GEORGE J. SUCI, 251

### NOTES:

*A Note on Functional Relations Obtained from Group Data:* MURRAY SIDMAN, 263

*A Note on the Combination of Ratings on the Basis of Reliability:* C. H. LAWSHE AND BRYANT F. NAGLE, 270

### BOOK REVIEWS:

TOLMAN's *Collected papers in psychology*: ROBERT B. MACLEOD, 274

KATONA's *Psychological analysis of economic behavior*: SAMUEL P. HAYES, JR., 276

ROHRER AND SHERIF's *Social psychology at the crossroads*: HAROLD GURTZKOW, 280

QUEENER's *Introduction to social psychology*: CLARENCE LEUBA, 282

MORENO's *Sociometry, experimental method and the science of society: An approach to a new political orientation*: JOHN HARDING, 283

VOLKART's *Social behavior and personality. Contributions of W. I. Thomas to theory and social research*: KIMBALL YOUNG, 284

BOHM's *Lehrbuch der Rorschach-Psychodiagnostik für Psychologen Ärzte und Pädagogen*: ZYGMUNT PIOTROWSKI, 287

BOOKS AND MATERIALS RECEIVED: 289

---

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

LYLE H. LANIER, Editor  
*University of Illinois*

LORRAINE BOUTHILET, Managing Editor

CONSULTING EDITORS

STUART H. BRITT  
*Needham, Louis and Brorby, Inc., Chicago*

DORWIN CARTWRIGHT  
*University of Michigan*

FRANK A. GELDARD  
*University of Virginia*

JAMES J. GIBSON  
*Cornell University*

DAVID A. GRANT  
*University of Wisconsin*

WILLIAM T. HERON  
*University of Minnesota*

ERNEST R. HILGARD  
*Stanford University*

WILLIAM A. HUNT  
*Northwestern University*

JEAN WALKER MACFARLANE  
*University of California*

DONALD G. MARQUIS  
*University of Michigan*

JOHN T. MCCALL  
*University of Vermont*

JAMES G. MILLER  
*University of Chicago*

NEAL E. MILLER  
*Yale University*

HELEN PARKS  
*University of Michigan*

ROBERT R. SEARS  
*Harvard University*

ROBERT L. THORNDIKE  
*Teachers College, Columbia University*

*The Psychological Bulletin* contains evaluative reviews of the literature in various fields of psychology; methodological articles, critical notes, and book reviews. This JOURNAL does not publish reports of original research or original theoretical articles.

Editorial communications concerning items to be printed during 1952 should be sent to Lyle H. Lanier, Department of Psychology, University of Illinois, Urbana, Illinois. All new manuscripts should henceforth be sent to the incoming editor, Wayne Dennis, Department of Psychology, Brooklyn College, Brooklyn, New York. Books for review should be sent to the new book review editor, Edward Girden, at the same address.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the article by Anderson and Valentine, "The preparation of articles for publication in the journals of the American Psychological Association" (*Psychological Bulletin*, 1944, 50, 345-376). Special attention should be given to the section on the preparation of the bibliography (pp. 363-372), since this is a particular source of difficulty in the reviews of research literature. All copy must be double-spaced, including the bibliography.

Reprints. Authors may order reprints when returning proof. Five copies of the JOURNAL are supplied gratis to contributors of articles, notes, special reviews, and book reviews. No reprints are supplied gratis.

Business communications—including subscriptions, orders of back issues and changes of address—should be sent to the American Psychological Association, 1515 Massachusetts Avenue, N. W., Washington 5, D. C.

Annual subscription: \$7.00 (Foreign \$7.50). Single copies, \$1.25.

PUBLISHED BI-MONTHLY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1515 Massachusetts Ave., N.W., Washington 5, D.C.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Accepted for mailing at special rate of postage provided for in Section 538, act of February 25, 1925, authorized January 2, 1947. Printed in U.S.A.

Copyright, 1952, by The American Psychological Association, Inc.

# Psychological Bulletin

## THE NATURE AND MEASUREMENT OF MEANING<sup>1</sup>

CHARLES E. OSGOOD

*University of Illinois<sup>2</sup>*

The language process within an individual may be viewed as a more or less continuous interaction between two parallel systems of behavioral organization: sequences of central events ("ideas") and sequences of instrumental skills, vocalic, gestural, or orthographic, which constitute the communicative product. A communicator vocalizes, "It looks like rain today; I'd better not wash the car." This output is a sequence of skilled movements, complicated to be sure, but not different in kind from tying one's shoes. Even the smallest units of the product, phonetic elements like the initial "l"-sound of "looks," result from precisely patterned muscle movements. The organization of these movements into word-units represents skill sequences of relatively high predictability; certain longer period sequences involving syntactical order are also relatively predictable for a given language system. But execution of such sequences brings the communicator repeatedly to what may be called "choice-points"—points where the next skill sequence is not highly predictable from the objective communicative product itself. The dependence of "I'd better not wash the car" upon "looks like rain today," the *content* of the message, reflects determinants within the semantic system which effectively "load" the transitional probabilities at these choice-points.

It is the communicative product, the spoken or written words which follow one another in varying orders, that we typically observe. Since we are unable to specify the stimuli which evoke these communicating reactions—since it is "emitted" rather than "elicited" behavior in Skinner's terminology (97)—measurements in terms of rates of occurrence and transitional probabilities (dependence of one event in the stream upon others) are particularly appropriate (cf., Miller, 76). In-

<sup>1</sup> The research on which in part this report is based is being supported jointly by the University Research Board of the University of Illinois and the Social Science Research Council. Grateful acknowledgment is made to both institutions for their assistance.

<sup>2</sup> Institute of Communications Research and Department of Psychology.

terest may be restricted to the lawfulness of sequences in the observable communicative product itself, without regard to the semantic parallel. This is traditionally the field of the linguist, but even here it has proved necessary to make some assumptions about meaning (cf., Bloomfield, 4). On the other hand, one may be specifically interested in the semantic or ideational level. Since he is presently unable to observe this level of behavior directly, he must use observable characteristics of the communicative product as a basis for making inferences about what is going on at the semantic level. He may use sequential orderliness in the product to draw conclusions as to semantic orderliness in the speaker's or writer's mediation processes (i.e., which "ideas" tend to go together in his thinking with greater than chance probabilities). Or he may wish to study the ways in which central, semantic processes vary from concept to concept, from person to person, and so on. It is the problem of measuring meaning in this latter sense which will be discussed in the present paper.

Before inquiring into the measurement of the meaning of signs, for which there are no accepted, standardized techniques available, we may briefly mention certain fairly standard methods for measuring the comparative strength of verbal habits. Thorndike and his associates (102, 103) have made extensive *frequency-of-usage counts* of words in English; that this method gets at the comparative habit strengths of word skill sequences is shown by the fact that other measures of response strength, such as latency and probability within the individual (Thumb and Marbe, 106; Cason and Cason, 19), are correlated with frequency-of-usage. Zipf (117, 118, and elsewhere) has described innumerable instances of the lawfulness of such habit-strength measures. Whether samples be taken from Plautine Latin, newspaper English, or the English of James Joyce in his *Ulysses*, a fundamental regularity is found, such that frequency of occurrence of particular words bears a linear relation to their rank order in frequency, when plotted on double-log paper (Zipf's Law). Measurement of flexibility or diversity in communicative products is given by the *type-token ratio* (TTR): with each instance of any word counting as a token and each different word as a type, the greater the ratio of types to tokens the more varied is the content of a message. This measure can be applied comparatively to different forms of material, different kinds of individuals, and so forth (cf., Carroll, 16, 17; Johnson, 45; Chotlos, 20), provided the sizes of samples are constant. One may also count the ratios of adjectives to verbs (Boder, 5), the frequencies of different pronouns, intensives, and so forth (cf., Johnson, 45).



Although the above measures get at the comparative strengths of verbal skill sequences per se (i.e., without regard to meaning), this is not a necessary restriction. Frequency counts of this type can be applied to *semantic habit strengths* as well. Skinner (96) has shown that a similar lawfulness applies to the frequencies of "free" associations in the Kent-Rosanoff tests. When frequencies of particular associates to given stimulus words for a group of subjects are plotted against their rank order in frequency, a straight-line function on double-log paper results (Zipf's Law). In other words, associations at the semantic level appear to be organized in such a way that few have very high probability of occurrence and many have low probabilities of occurrence. Bousfield and his collaborators (7, 8, 9, 10, 11) have described a *sequential association method* for getting at comparative semantic habit strengths. When subjects associate successively from the same "pool," e. g., "names of four-legged animals," (a) the rate of successive associates shows a negatively accelerated curve, (b) varying in its constants with certain characteristics of materials and subjects, (c) the order of appearance of particular associates in individuals being predictable from the frequency of usage in the group, and (d) distortions in the function being related to particular transitional probabilities among associates, i.e., clustering. Useful though these measures are for many purposes, they do not get at meaning. The fact that "dog" has a higher probability of occurrence in sequential association than "otter" says nothing whatsoever about the differences in meaning of these two signs.

An extensive survey of the literature fails to uncover any generally accepted, standardized method for measuring meaning. Perhaps it is because of the philosophical haziness of this concept, perhaps because of the general belief that "meanings" are infinitely and uniquely variable, or perhaps because the word "meaning" as a construct in our language connotes mental stuff, more akin to "thought" and "soul" than to anything observable—for some combination of reasons there has been little attempt to devise methods here. Nevertheless, whether looked at from the viewpoints of philosophy or linguistics, from economic or sociological theory, or—interestingly enough—from within the core of psychological theories of individual behavior, the nature of meaning and change in meaning are found to be central issues. The proposals to be made in the latter portion of this paper are part of a program aimed at the development of objective methods of measuring meaning. Beyond obvious social implications, it is felt that this direction of research is a logical extension of scientific inquiry into an area generally considered immune to its attack.

## THEORIES OF MEANING

Not all stimuli are signs. The shock which galvanizes a rat into vigorous escape movements usually does not stand for anything other than itself, nor does the pellet of food found at the end of a maze, nor a hammer in one's hand or a shoe on one's foot. The problem for any meaning theorist is to differentiate the conditions under which a pattern of stimulation is a sign of something else from those conditions where it is not. This certainly seems simple enough, yet it has troubled philosophers for centuries. By stating the problem somewhat formally, the chief differences between several conceptions of the sign-process can be made evident: let

$\dot{S}$  = object = any pattern of stimulation which evokes reactions on the part of an organism, and

$\square$  = sign = any pattern of stimulation which is not this  $\dot{S}$  but yet evokes reactions relevant to  $\dot{S}$ —conditions under which this holds being the problem for theory.

The definition of  $\dot{S}$  is broad enough to include any pattern of stimulation which elicits any reaction from an organism. Although one usually thinks of "objects" as those things denoted by signs, actually any pattern of stimulation—a gust of northerly wind against the face, the sensations we call "belly-ache," the sensations of being rained upon—is an "object" at this level of discourse. One sign may be the "object" represented by another sign, as when the picture of an apple is called "DAX" in certain experiments. The definition of  $\square$  is purposely left incomplete at this point, since it depends upon one's conception of the nature of the sign-process.

We may start a logical analysis of the problem with a self-evident fact: *the pattern of stimulation which is the sign is never identical with the pattern of stimulation which is the object.* The word "hammer" is not the same stimulus as is the object hammer. The former is a pattern of sound waves having characteristic oscillations in frequency and intensity; the latter, depending upon its mode of contact, may be a visual form having characteristic color and shape, a pattern of tactual and proprioceptive sensations, and so on. Similarly, the buzzer in a typical rat experiment is not identical as a form of stimulation with the shock which it comes to signify. Yet these signs—the word "hammer" and the buzzer—do elicit behaviors which are in some manner relevant to the objects they signify, a characteristic *not* shared with an infinite number of other stimulus patterns that are *not* signs of these objects. In simplest terms, therefore, the question is: *under what conditions does something*

which is not an object become a sign of that object? According to the way in which this question is answered we may distinguish several theories of meaning.

### *Mentalistic View*

The classic interpretation derives directly from the natural philosophy of Western culture, in which the dualistic connotations of language dictate a correlation between two classes of events, material and nonmaterial. Since meanings are obviously "mental" events and the stimuli representing objects and signs are obviously "physical" events, any satisfying theory of meaning must specify interrelation between these levels of discourse. At the core of all mentalistic views, therefore, we find an "idea" as the essence of meaning; it is this mental event which links or relates the two different physical events, sign and object. The word "hammer" gives rise to the idea of that object in the mind; conversely, perception of the object hammer gives rise to the same idea, which can then be "expressed" in appropriate signs. In other words, *something which is not the object becomes a sign of that object when it gives rise to the idea associated with that object*. Probably the most sophisticated expression of this view is given by Ogden and Richards (82) in their book, *The Meaning of Meaning*. Most readers will recall their triangular diagram of the sign-process: the relation between symbol and referent (the base of their triangle) is not direct but inferred, mediated through mental "thought" or "interpretation" (the third corner of their triangle).

### *Substitution View*

Naive application of Pavlovian conditioning principles by early behaviorists like Watson led to the theory that signs achieve their meanings simply by being conditioned to the same reactions originally made to objects. This, in essence, is the view one encounters in many introductory texts in general psychology. An object evokes certain behavior in an organism; if another pattern of stimulation is consistently paired with the original object, it becomes conditioned to the same responses and thus gets its meaning. The object is the unconditioned stimulus and the sign is the conditioned stimulus, the latter merely being substituted for the former. The definition of the sign-process here is that *whenever something which is not the object evokes in an organism the same reactions evoked by the object, it is a sign of that object*. The very simplicity of this theory highlights its inadequacy. Signs almost never evoke the *same* overt responses as do the objects they represent. The word FIRE has meaning to the reader without sending him into headlong flight.<sup>9</sup> Nevertheless, this represents a first step toward a behavioral interpretation of the sign-process.

*Meaning as "Set" or "Disposition"*

In a monograph entitled *Foundation of the Theory of Signs* (77), Charles Morris, a semiotician working in the tradition established by Peirce and other American pragmatists, proposed a formula for the sign-process which avoids the pitfalls of substitution theory but seems to step backward toward the mentalistic view. In essence he states that signs achieve their meanings by eliciting reactions which "take account of" the objects signified. The sign "hammer" may evoke quite different responses from those evoked by the object signified, but these responses must have the character of being relevant to the object. The response made to the sign is called the "interpretant" which mediately takes account of the object signified. But it would seem that this process of "taking account of" is precisely what needs elucidation.

During the period intervening between this monograph and his recent book, *Signs, Language and Behavior* (78), Morris studied with two prominent behavior theorists, Tolman and Hull. The effects of this immersion in learning theory are evident in his book, which is a pioneer attempt to reduce semiotic to an objective behavioral basis. He states that "if anything, *A*, is a preparatory stimulus which in the absence of stimulus-objects initiating response-sequences of a certain behavior-family causes a disposition in some organism to respond under certain conditions by response-sequences of this behavior-family, then *A* is a sign" (p. 10). Reduced to its essentials and translated into our terms, this becomes: *any pattern of stimulation which is not the object becomes a sign of that object if it produces in an organism a "disposition" to make any of the responses previously elicited by that object.* There is no requirement that the overt reactions originally elicited by the object also be made to the sign; the sign merely creates a disposition or set to make such reactions, actual occurrence depending upon the concurrence of supporting conditions.

Beyond the danger that "dispositions" may serve as mere surrogates for "ideas" in this theory, there are certain other difficulties with the view as stated. For one thing, Morris seems to have revived the substitution notion. The sign is said to dispose the organism to make overt response-sequences of the *same* behavior-family originally elicited by the object. But is this necessarily the case? Is my response to the word "apple" (e.g., free-associating the word "peach") any part of the behavior-family elicited by the object apple? For another thing, Morris' formulation fails to differentiate sign-behavior from many instinctive reactions and from ordinary conditioning. To appreciate this difficulty will require a brief digression.

When a breach is made in a termite nest, the workers set up a distinctive pounding upon the floor of the tunnel and the warriors come charging to the spot, where they take up defensive positions. Is this pounding sound a sign to the warrior-termites that there is a breach in the nest? It happens that this



behavior is purely instinctive, and most students of sign-behavior believe that signs must achieve their signification through *learning*. But is learning a sufficient criterion? Are all stimuli that elicit learned reactions automatically signs? In developing any skill, such as tying the shoes, the proprioceptive stimuli produced by one response become conditioned to the succeeding response—but of what are these proprioceptive stimuli signs? With repeated experience on an electrified grill a rat will often learn to rear up on its hind legs and alternately lift them, this act apparently reducing the total intensity of pain—the painful stimulation is thus conditioned to a new response, but of what is the pain a sign?

If only some of the stimuli which elicit learned responses are signs, we must seek a reasonable distinction *within* the class of learned behaviors. We cannot draw a line between human and subhuman learning: the buzzer is operationally as much a sign of shock to the rat in avoidance-training experiments as are dark clouds a sign of rain to the professor—both stimuli elicit reactions appropriate, not to themselves, but to something other than themselves. Is voluntariness of response a criterion? Meaningful reactions may be just as involuntary as perceptions—try to observe a familiar word and avoid its meaning! Is it variability of response to the stimulus? Meaningful reactions may be just as stable and habitual as motor skills.

### *The Mediation Hypothesis*

I shall try to show that the distinguishing condition of sign behavior is the presence or absence of a *representational mediation process* in association with the stimulus. This conception of sign behavior is based upon a general theory of learning rather than being concocted specifically to account for meaning as seen in human communication.<sup>3</sup> The essence of the viewpoint can be given as follows:

1. *Stimulus-objects (S) elicit a complex pattern of reactions from the organism, these reactions varying in their dependence upon presence of the stimulus-object for their occurrence.* Electric shock galvanizes the rat into vigorous jumping, squeaking, and running activities, as well as autonomic "anxiety" reactions. Food objects elicit sequences of salivating, chewing, lip-smacking, and so forth. Components like salivating and "anxiety" are relatively independent of the food or shock stimulation respectively and hence can occur when such objects are not present.

2. *When stimuli other than the stimulus-object, but previously associated with it, are later presented without its support, they tend to elicit some reduced portion of the total behavior elicited by the stimulus-object.* This reduction process follows certain laws: (a) mediating reactions which interfere with goal-achievement tend to extinguish; (b) the more energy expenditure involved in making a particular reaction, the less likely it is to survive the reduction process; (c) there is evidence that certain reactions (e.g., autonomic) condition more readily than

<sup>3</sup> This hypothesis, as an elaboration from Hullian theory (43), is described in my forthcoming book, *Method and Theory in Experimental Psychology*.



others (e.g., gross skeletal) and hence are more likely to become part of the mediation process—this may merely reflect factor (b) above.

3. *The fraction of the total object-elicited behavior which finally constitutes the stable mediation process elicited by a sign ( $\Sigma$ ) will tend toward a minimum set by the discriminatory capacity of the organism.* This is because the sole function of such mediating reactions in behavior is to provide a distinctive pattern of self-stimulation (cf., Hull's conception of the "pure stimulus act").

4. *The self-stimulation produced by sign-elicited mediation processes becomes conditioned in varying strengths to the initial responses in hierarchies of instrumental skill sequences.* This mediated self-stimulation is assumed to provide the "way of perceiving" signs or their "meaning," as well as mediating instrumental skill sequences—behaviors to signs which take account of the objects represented.

Whereas Morris linked sign and object through partial identity of object-produced and disposition-produced behaviors, we have linked sign and object through partial identity of the "disposition" *itself* with the behavior elicited by the object. Words represent things because they produce some replica of the actual behavior toward these things, as a mediation process. This is the crucial identification, the mechanism that ties particular signs to particular stimulus-objects and not to others. Stating the proposition formally: *a pattern of stimulation which is not the object is a sign of the object if it evokes in an organism a mediating reaction, this (a) being some fractional part of the total behavior elicited by the object and (b) producing distinctive self-stimulation that mediates responses which would not occur without the previous association of nonobject and object patterns of stimulation.* This definition may be cumbersome, but all the limiting conditions seem necessary. The mediation process must include some part of the same behavior made to the object if the sign is to have its particularistic representing property. What we have done here, in a sense, is to make explicit what may be implicit in Morris' term "disposition." The second stipulation (b) adds the learning requirement—the response of warrior-termites to pounding on the tunnel floor is ruled out since it does not depend upon prior association of pounding with discovery of a breach in the nest.

Paradigm A in Figure 1 gives an abbreviated symbolic account of the development of a sign, according to the mediation hypothesis. Take for illustration the connotative meaning of the word SPIDER. The stimulus-object ( $\dot{S}$ ), the visual pattern of hairy-legged insect body often encountered in a threat context provided by other humans, elicits a complex pattern of behavior ( $R_T$ ), which in this case includes a heavy loading of autonomic "fear" activity. Portions of this total behavior to the spider-object become conditioned to the heard word, SPIDER. With repetitions of the sign sequence, the mediation process becomes reduced to some minimally effortful and minimally interfering replica—but still includes those autonomic reactions which confer a threatening

significance upon this sign. This mediating reaction ( $r_m$ ) produces a distinctive pattern of self-stimulation ( $s_m$ ) which may elicit a variety of overt behaviors ( $R_x$ )—shivering and saying "ugh," running out of a room where a spider is said to be lurking, and even refusing a job in the South, which is said to abound in spiders.

The vast majority of signs used in ordinary communication are what we may term *assigns*—their meanings are literally "assigned" to

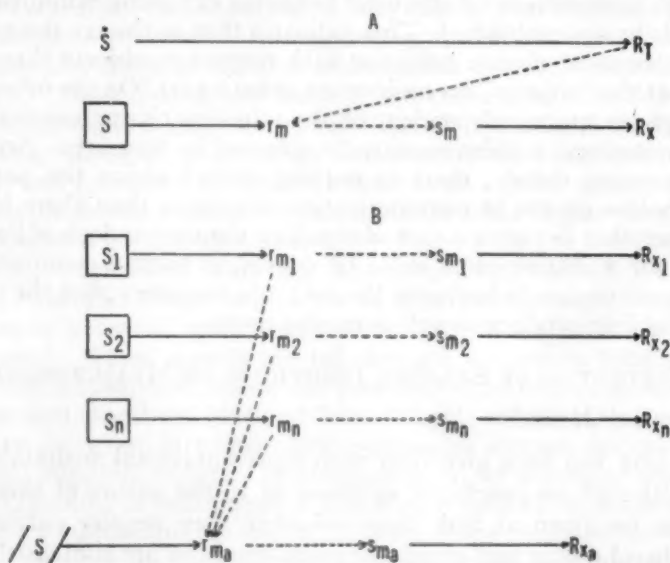


FIG. 1. SYMBOLIC ACCOUNT OF THE DEVELOPMENT OF SIGN PROCESSES:

A. DEVELOPMENT OF A SIGN; B. DEVELOPMENT OF AN ASSIGN.

them via association with other signs rather than via direct association with the objects represented. The word ZEBRA is understood by most six-year-olds, yet few of them have ever encountered zebra-objects themselves. They have seen pictures of them, been told they have stripes, run like horses, and are usually found wild. As indicated in Figure 1 (B), this new stimulus pattern, ZEBRA, "picks up" by the mechanisms already described portions of the mediating reactions already elicited by the primary signs. In learning to read, for example, the "little black bugs" on the printed page are definitely assigns; these visual patterns are seldom directly associated with the objects signified, but rather with auditory signs (created by the child and teacher as they verbalize). Obviously, the more quickly the child can learn to make the right noises to these visual stimuli (the modern phonetic approach to reading), the more quickly these new, visual assigns will acquire signifi-

cance. The child already has meanings for HOUSE, DOG, and even TYPEWRITER as *heard* stimulus patterns, but these mediation processes must be assigned to *seen* stimulus patterns.

It is apparent from the foregoing that the meanings which different individuals have for the same signs will vary with their behaviors toward the objects represented. This is because the composition of the mediation process, which is the meaning of a sign, is entirely dependent upon the composition of the total behavior occurring while the sign-process is being established. This indicates that to change the meaning of signs we must change behavior with respect to objects (keeping in mind that the "objects" for assigns are other signs). On the other hand, meanings are quite independent of the stimulus characteristics of the signs themselves, a point repeatedly stressed by linguists. According to the present theory, there is nothing sacred about the particular mouth-noises we use in communication any more than there is about the buzzer that becomes a sign of shock to the rat—a flash of light or a blast of air would serve as well. Of course, in human communication (in contrast to sign-behavior in the rat) it is necessary that the users of signs be able to produce as well as receive them.

#### EVALUATION OF EXISTING TECHNIQUES OF MEASUREMENT

##### *Physiological Methods*

Meaning has been identified with representational mediation processes. Although no conclusive evidence as to the nature of these processes can be given at this time—whether they require participation of peripheral motor and glandular mechanisms or are sufficiently characterized as central phenomena—it is convenient to conceive of them as implicit response processes which produce self-stimulation. The meager evidence available certainly does not refute this view. In any case, the investigator is encouraged to see whether or not any physiological measures display covariance with changes in meaning. Correlates of this sort would be direct indices of meaning.

*Action potentials in striate musculature.* Working under the impetus given by the Würzburg School of imageless thought, introspective psychologists of another generation tried to tease out the sensory content of ideas. Students like Marbe (69), Clark (21), Washburn (110), Comstock (25) and Crosland (26) agreed in finding kinaesthetic sensations present as a residue when everything but "meaningful thought" was excluded. But here the limitations inherent in the introspective method made themselves apparent: pressed to the limits of critical self-analysis, the trained human observer reported vague muscular and organic sensations as being present during thought—but did these sensations constitute thoughts and meanings themselves or were they merely a back-

ground of bodily tonus? The method did not permit this kind of discrimination.

Experimentalists picked up the problem at precisely the point where the introspectionists had per force dropped it—was it possible that sufficiently sensitive apparatus could record the minimal motor events that eluded conscious analysis? Watson's (111) statement about thought being implicit speech was the "open sesame" for a small host of gadgeteers (Wyczoikowski, 116; Reed, 89; Clark, 22; Thorson, 104) who filled subjects' mouths with an astounding variety of mechanical devices and then had them both think and mumble unusual items like "psychology." That they found little or no correspondence between the movements made during thought and speech is not particularly surprising. The "thought" movements were results of irritation in all probability, and (contrary to Watson) there is no a priori reason to expect relevant mediation processes to be restricted to the organs of speech.

Apparatus of sufficient sensitivity came with the development of electrical methods of recording and amplifying action. Electrodes placed near the motor end-plates of efferent fibers pick up minute changes in potential which cause deflections in a sensitive string galvanometer. These signals are fed through a vacuum tube amplifier, magnified thousands of times, and recorded on photographic film. Using this method, Jacobsen (44) and Max (72, 73) obtained suggestive correlations between introspectively specifiable events and objectively recorded muscular events. Jacobsen showed, for example, that when a subject, trained in techniques of progressive relaxation, *imagines* lifting his right arm, distinctive potential changes are recorded from muscles in that member but not elsewhere. Max, using deaf-mutes as an ingenious control, was able to show that these individuals display more frequent and larger potential changes in the muscles controlling their fingers while solving mental problems and while dreaming than normal individuals; he also described a negative correlation between the magnitude of such implicit activity and intelligence scores, i.e., more intelligent subjects showed less overt activity.

Are these recordable motor events the mediating reactions ( $r_m$ ) in the mediation processes we have specified as essential in sign behavior? Does this technique provide an index for the presence or absence of meaning, its degree, and quality? This is an attractive possibility, but the index is a crude one. There is no way of "reading" the meaning of a sign to a subject from the recorded activity. The ultimate criterion of meaning is still introspection of the subject—he verbalizes meaning while the experimenter scurries about his periphery trying to pick it up on instruments, and when activity fails to appear at predicted locus *a* it is assumed it must be occurring at some other locus *b*. No satisfactory demonstration of the *necessity* of the motor component has been offered; meaning might be present without measurable motor activity which,



when it does occur, is simply a specific overflow of excitation into motor pathways—an epiphenomenon. The same criticism applies with equal force to the other physiological correlates that have been studied.

*Salivary reaction.* Another pioneer investigation into the organic correlates of meaning was that by Razran (87), serving as his own subject. Meaningfulness of a series of signs was the independent variable, the stimuli being words for "saliva" in languages with which Razran had varying familiarity. Amount of salivary secretion was the dependent variable—following presentation of each stimulus, a dental cotton-roll was placed in his mouth for two minutes and its weight determined immediately afterward. As "meaningless" controls he used the Gaelic word for saliva, the nonsense syllables QER SUH, and periods of "blank consciousness." Salivation was greatest in his childhood tongue (Russian), next in his most proficient one (English), and less in three slightly known languages (French, Spanish, and Polish). The control conditions showed no differences among themselves, despite the fact that Razran "knew" the Gaelic word stood for saliva. This experiment demonstrates a relation between amount of salivation and degree of meaningfulness of signs to a sophisticated subject. We have here another feasible index of some aspect of meaning, albeit a very limited one.

*The galvanic skin response.* The GSR is one of several indices of autonomic activity, and to the extent that meanings include emotional components this measure should be useful. There are a large number of studies using the GSR that are remotely related to this problem: GSR is readily elicited by any warning or preparatory stimulus that precedes shock (Darrow and Heath, 28; Switzer, 101; Mowrer, 80), and it, therefore, may serve as an indication that the preparatory stimulus has become a sign of shock. It has been used to index the intensity of pleasant or unpleasant connotations of words and experiences (Jones, 46; Lynch, 63). In connection with free association, GSR has been found to be a good indicator of the emotional effect of stimulus words (Jones and Wechsler, 47). It is unfortunate, therefore, that the two most pertinent experiments in this area leave much to be desired in the way of methodological finesse.

Mason (70) asked this question: do changes in GSR accompany changes in meaning? What were called three types of change in meaning were studied: "certainty of meaning" (the expression of certainty by the subject as to the correctness of his recall in learning a list of nonsense syllables), "discovery of meaning" (the point in a series of readings of a trick sentence without punctuation at which the subject achieved insight), and "loss of meaning" (where the subject pressed a foot-pedal whenever the continuously vocalized word "tangerine" seemed to lose its meaning). Although the procedures and results of this experiment are reported in great detail, no tests of significance were employed and none of the necessary controls was introduced. In the



first two experiments, for example, it is impossible to determine whether the deflections in GSR should be attributed to change in meaning or simply to vacillations in emotional stress (the latter seems quite likely). In the experiment on loss of meaning, no check was made to see what effect simply pressing the indicator pedal and seeing a signal light come on might have had upon GSR.

Bingham (3) measured psychogalvanic reaction to 72 words "selected from the educational and philosophical writings of John Dewey and Rabindranath Tagore as the most frequently occurring words in samples of 8,000 from each writer." After the galvanic measurements were made, the 50 undergraduate subjects rated the words on a three-point scale in terms of their personal meaningfulness, significance, and importance (combined into a single scale, MSI). These untrained subjects further introspected on the sensory content in the meaning of each of these words, being requested to report any visual, auditory, kin-aesthetic, etc., imagery or sensations. Words having the highest MSI ratings yielded the greatest average change in skin resistance. High MSI words also had more "organic" sensory content, according to the introspections (the high "organic" content words were *intellectual, freedom, God, truth, and love*, in this order!). Much remains to be done in these directions.

### *Learning Methods*

There are many learning studies employing meaningful materials, but rarely is meaning itself the experimental variable. Even where meaning has been deliberately varied, interest has generally centered on the effect upon learning rather than upon the use of learning as an index of meaning. Only the more relevant studies will be considered here.

*Semantic generalization.* When a reaction conditioned to one stimulus transfers to another, and the amount of transfer varies directly with the similarity between the two stimuli, we speak of stimulus generalization. The operations whereby semantic generalization is demonstrated are the same—except that the necessary similarities lie in meaning rather than in objective physical characteristics of the eliciting stimuli. There is no physical similarity between the word "blue" and light of 450  $m\mu$  yet generalization between such stimulations is easily demonstrated—we infer that the common overt reaction is mediated by some common implicit process. In the experiments to be summarized here, the precise nature of the overt reaction is unimportant—all the standard CR's have been used, salivation, GSR, finger retraction, pupillary reflex, and so on. Much of the research in this area is contributed by Russian investigators and, unfortunately, the available reports are mostly in the form of brief abstracts.

1. *From object to sign.* Kapustnik (49) set up conditioned salivary reactions to visual and auditory stimuli, transfer to verbal signs for the original cues being tested. Kotliarevsky (59) employed a cardio-vasomotor reflex: following conditioning to the sound of a bell, response to the word "bell" was tested. Metzner (75) reports a similar experiment with the pupillary reflex. In all these cases significant amounts of generalization were obtained. Traugott (107) was able to demonstrate the generalization of conditioned inhibition from blue light to the words, "blue, blue"—in fact, these words (quite different as physical stimuli) showed greater generalization than actual red light. This investigator was also able to show that the inhibitory effect of the words "blue, blue" transferred to other conditioned reactions more broadly than did the effect of blue light itself, a finding which fits well with our notions as to the abstractability of sign-processes. Traugott and Fadeyeva (108) combined conditioning and free association techniques: with excitatory CR's set up to a bell-whistle-light pattern and inhibitory CR's to a whistle-touch pattern, free associations to the verbal signs of these stimuli were recorded along with the latencies of these associations. Association to words representing conditioned excitors were made more rapidly than those to words representing conditioned inhibitors and, interestingly enough, after extinction of the excitatory CR's associations became slower and generically older, i.e., associations which had referred to the experimental situation now referred to pre-experimental situations.

2. *From sign to sign.* When a response is conditioned to one sign (e.g., the word TREE) and generalization to other signs (e.g., BUSH or the picture of a tree or bush) is measured, the essential role of meaningful mediation is merely more obvious than in the preceding situation. Razran (88) flashed single words and short sentences on a screen while six adults were eating—conditioned salivation developed rapidly. In a second session, different words and sentences were used to measure generalization. For the single words generalization was found to be greater for semantically related words (e.g., STYLE to FASHION) than for phonetically related words (e.g., STYLE to STILE), a result in keeping with Traugott's results discussed above and data more recently obtained by Riess (90) using the GSR as a measure. A slightly discordant note is contributed by Keller (55). After conditioning the GSR to a picture of a boy-scout hat, tests for generalization were made for a picture of a fireman's hat and the printed word HAT, neutral control items being pictures of a duck and baseball and the words DUCK and BALL. While significant generalization to the picture of a fireman's hat occurred, no transfer to the word HAT was obtained. Keller argues reasonably that if the generalization between the two pictured hats was based on a common mediating response, thinking or subvocalizing "hat," the printed word should have shown the same effect. In one of the most interesting studies of this type, Riess (91) has related semantic generalization to stages in genetic development. Four groups of subjects, varying in mean age from 7:9 years to 18:6 years of age, were trained to give the GSR to selected verbal stimuli. Tests for transfer to synonyms, antonyms, and homonyms of the original words were run. The generalization results indicate that meaningful or semantic similarities (synonym and antonym relations) increase in importance as the individual matures while the importance of physical similarities (homonym relation) decreases.

3. There are other semantic relations that could be studied with similar

techniques, but there is little evidence available. Generalization between *hierarchical levels of signs*, e.g., between DOG and ANIMAL, has been studied by Goodwin, Long, and Welch (36), but there are certain difficulties with the research design (see below). Generalization would also be expected to occur *from sign to object*, even though this seems to reverse the sequence followed in the development of meaning. Kapustnik (49) found that salivary reactions conditioned to verbal signs transfer to the stimuli signified. This is the only directly relevant study I have uncovered, but observations on much of social behavior fit the paradigm. The prejudicial reactions associated with "Wop" and "Jap" on a verbal stereotype level certainly tend to transfer to the social objects represented, once they are encountered. Finally, mention should be made of generalization *from object to object via semantic mediation*. The reverential care with which the adolescent handles a certain handkerchief, a certain lock of hair, a certain lipstick-printed napkin has nothing to do with physical similarities among these objects themselves. Similarly, an "inferiority complex" may render a wide range of physically dissimilar social objects and situations equivalent in meaning and hence reaction to a particular individual (cf., G. W. Allport's [1] trait hypothesis).

Cofer and Foley have related the various studies of semantic generalization to the theoretical mediation process. They state that semantic generalization "thus presupposes and depends upon the preexperimental formation of conditioned responses or associations, i.e., *the gradient of generalization is a gradient along a dimension of conditioned stimulus functions*. The stimuli need be similar only in so far as they have previously been conditioned to the same (or similar) response" (23, p. 520). The pre-experimental formation of conditioned responses to which they refer is a special case of the formation of mediation processes, as discussed earlier in this paper. It is probably necessary to assume that primary generalization occurs both among mediating reactions and the stimuli they produce, in order to account for the fact that *gradients* of semantic generalization are correlated with *degrees* of meaningful similarity.

*Transfer and interference in learning.* The experiments contributed by Cofer and Foley in support of their hypotheses fit the standard transfer design. The general procedure was as follows: first a single repetition on a *buffer list* of numbers (spelled out) was given as warm-up; then an *equating list* of proper names was presented once and scored for recall immediately, subjects being assigned to various experimental conditions on this basis; each subject was then given four unscored repetitions on either a *reinforcement list* or a *control list*; finally, all subjects were tested for recall of a *test list* of words immediately after a single presentation. Experimental, control and test lists used in the first experiment (Foley and Cofer, 32) are given in Table 1. All words on a given reinforcement list bear the same relation to the test list, either some degree of synonymy or some degree of homonymy.

Although all experimental conditions yielded better recall scores than the control condition, there are several curious points about these results. In the first place, the best homonym list shows more "generalization" to the test list than the best synonym list—this is in flat contra-

diction to most other findings (cf., Traugott, Razran, Riess above). Secondly, the difference in recall between the two homonym lists is clearly greater than that between the two synonym lists, yet the former are obviously equivalent (both being identical in sound to the test words) while the latter are definitely not equivalent (the words on I are close synonyms of the test words but those on II bear no relation whatever to the test words, viz., sent-killed, vein-help, pear-result, sow-factory). Fortunately, the difference between "Synonym II" and the control condition was not significant.

TABLE 1  
MATERIALS FOR AN EXPERIMENT ON MEDIATED TRANSFER  
(FOLEY AND COFER, 32)

Control List	Reinforcement Lists				Test List
	Homonym I	Homonym II	Synonym I	Synonym II	
palm	cent	scent	dispatched	killed	sent
set	vain	vane	vessel	ship	vein
reed	pare	pair	fruit	result	pear
very	sew	so	plant	factory	sow
numb	rain	rein	rule	principle	reign
me	seas	seize	looks	appearance	sees
day	write	rite	just	barely	right
snap	noes	nose	apprehends	arrests	knows
rope	meat	mete	join	enlist	meet
spire	dō	doe	batter	bruise	dough
Mean Number of Words on Test List Recalled Following Above					
4.80	6.72	5.64	5.88	5.24	

In searching for an explanation of these points, a flaw in design was discovered which renders this entire technique suspect. Since all the words on a given reinforcement list bore the same relation to those on the test list, all a subject had to do was to "catch on" to this abstract relation and then proceed to manufacture the test list rather than recall it. Given four trials on a list of only 10 meaningful words, the subject presumably masters most of it. If, on the single presentation of the test list, he now notes that the new words are homonyms of the old, by merely recalling *cent*, *vain*, and *pare* he can do a pretty good job of manufacturing (and checking by recognition) *sent*, *vein*, and *pear*. He can do the same thing with synonym lists, but here he will make more errors since there are more alternatives. The same loophole in design is even more apparent in later studies in this series. In a study on *antonym gradients*, for example, Cofer, Janis, and Rowell (24) themselves point out that 19 of 28 subjects reported that they recognized the opposition relation. Foley and Mathews (34) and Goodwin, Long and



Welch (36) report experiments using the same method, and their results are of dubious value for the same reasons.

Two experiments on interference in verbal learning by Osgood (83, 84) were explicitly designed to test certain hypotheses regarding the nature of meaning. The following hypotheses were set up: (1) Words of opposed meaning are so because they elicit reciprocally antagonistic mediating processes. (2) Repeated reinforcements of the association between a new stimulus and a particular mediating reaction produce a negatively accelerated increase in the excitatory tendency associating this stimulus with this reaction *and simultaneously an equal inhibitory tendency associating this stimulus with the reciprocally antagonistic reaction*. In other words, in learning to make a reaction to a stimulus the organism is simultaneously learning *not* to make the directly antagonistic reaction to that stimulus. (3) Both excitatory and inhibitory tendencies generalize in the usual fashion among both mediating reactions and the stimuli they produce.

With nonsense letter-pairs as constant stimuli and meaningful adjectives as varied responses in the standard retroactive interference paradigm ( $A-B$ ;  $A-K$ ;  $A-B$ ), it was predicted that for both transfer and retroaction tests interpolated responses *similar* in meaning to the original responses should show the least interference (generalization of excitatory tendency) and responses *opposed* in meaning should show the most interference (generalization of inhibitory tendency), as compared with an intermediary neutral condition. The total design was such that each subject learned an equal number of items in each meaningful relation, thus avoiding the type of set that troubled the Foley and Cofer studies. The results for both transfer and retroaction situations were essentially those predicted (83). The second study (84) offered further evidence that a special form of reciprocal inhibition is operating in the successive learning of opposed meanings for the same sign. Different groups of subjects were given varying degrees of learning on the interpolated materials (cf. the design used by Melton and Irwin, 74) and only similar and opposed meaningful responses were compared. The learning of opposed responses was characterized by longer latencies of reaction and more frequent blanks (failures of response), both increasing with the degree of interpolated learning. Since both these phenomena are characteristic of weakened habit strength, they follow from the hypotheses.

These findings point to the following general conclusion: *when a sign or assign is conditioned to a mediator, it will also tend to elicit other mediators in proportion to their similarity to the original reaction; it will tend to inhibit other mediators in proportion to the directness of their antagonism to the original reaction*. In everyday language, this indicates that signs which develop a certain meaning through direct training will readily elicit similar meanings but resist being associated with opposed meanings. If the sign RUSSIAN means *bad* to the conservative college student he easily accepts substitution of *dirty*, *unfair*, and *cruel*, but it is



difficult for him to think of Russians as *clean*, *fair*, and *kind* (cf., Stagner and Osgood, 100).

### *Perception Methods*

There is an intimate relation between perceptual and meaningful phenomena (14, 105). It is borne out by the confusions psychologists display in using these terms. In one of Maier's (67) ingenious insight situations, for example, the crux of the problem lay in whether or not the human subject could shift from utilizing the handle of an ordinary lab clamp as something to tighten (original use) to something to hang one's hat on (use which would solve the problem). We could say that this handle must be "perceived differently" or the "field restructured perceptually" (cf., Köhler, 58; Wertheimer, 112), or that it must be "given a new functional value" (cf., Duncker, 30), or that it must "acquire a new meaning or significance as a stimulus." The voluminous literature on memory for forms has been interpreted both as demonstrating perceptual dynamics (cf., Koffka, 57) and semantic dynamics (cf., Bartlett, 2)—witness particularly the experiment by Carmichael, Hogan, and Walter (15) in which the deliberate introduction of different meaningful words in association with the same abstract forms markedly influenced the way they were recalled.

As was the case with the learning approach, there are few experiments in which meaning has been deliberately introduced as a variable. Remotely relevant are a group of studies in which the effect of hunger upon perception in ambiguous situations has been measured (Sanford, 93, 94; Levine, Chein, and Murphy, 62; McClelland and Atkinson, 64). The way of perceiving the ambiguous stimuli was clearly modulated by the presence or absence of this motive state. Postman and Bruner (86) have studied the effect of a different motive state, frustration, upon the perception of tachistoscopically presented sentences. Most significant from our present point of reference was the marked increase in "aggressive" and "escape" words as misperceptions following frustration. Generalizing, we might say that the internal *motive state* of the individual, as part of the total stimulus context, changes the probabilities of occurrence (availability) of alternate mediating processes for the same external stimulus. Whether the known *value* of an object, as one dimension of its meaning, can influence the way it is perceived is a moot question at this time. Bruner and Goodman (12), with apparent size of coins as the perceptual characteristic measured and rich or poor 10-year-olds as subjects, obtained what may be interpreted as positive results; Carter and Schooler (18), under generally similar conditions, failed to substantiate the earlier conclusions. Mausner and Siegel (71),

using recognition-time as a measure and stamps of varying value as stimuli, also report negative conclusions. Most relevant to the problem of measuring meaning are the following experiments.

Bruner and Postman (13) compared the apparent sizes of a dollar sign (positive symbol), swastika (negative symbol), and an abstract geometrical design (neutral control), as estimated by manipulating a spot of light when plastic discs of identical size bearing these symbols were held in the subject's hands. Both dollar sign and swastika showed significant *overestimation*. According to the investigators, two dynamic processes were operative: (a) perceptual enhancement due to the positive value of the dollar sign and (b) perceptual accentuation of apparent size due to the swastika alerting the organism to danger or threat—a single process of enhancement in size due to distinctiveness might be more parsimonious, and there are many other possible hypotheses.

Postman, Bruner, and McGinnies (86) hypothesized that personal values (as defined by scores on the Allport-Vernon test) are among the behavioral or attitudinal determinants of perception. Twenty-five subjects were shown 36 words, one at a time, in a modified Dodge tachistoscope, these words being presented in random orders and being chosen to represent the six Allport-Vernon values. The usual method for obtaining recognition thresholds was used—gradually increasing the flash exposure time until the subject correctly identifies the word. Take for illustration a subject with high social values and low theoretical: according to the investigators, his threshold for perceiving words like "loving" and "devoted" should be lowered by *selective sensitization*; in the presolution period he should misperceive words covalent with the correct word because of *value resonance*; and his threshold for words like "verify" and "research" should be raised because of *perceptual defense*. The results were consistent with the general thesis, but the single principle of selective sensitization seems sufficient to explain them.

McGinnies (65) inquired more penetratingly into the matter of "perceptual defense," conjecturing that autonomic reactions are aroused prior to conscious awareness of the meaning of a threatening word and hinder its perception. A list of 11 neutral words and seven emotionally charged words were presented tachistoscopically and recognition thresholds determined in the usual manner; GSR was recorded as a measure of emotional disturbance. Taboo words were found to require longer exposures for recognition and their prerecognition presentations were accompanied by significantly stronger emotional reactions. When asked if they had reported their perceptions promptly and accurately, all undergraduate subjects said they had.

Both this study and the preceding one on values have come in for their share of criticism. Howes and Solomon (40) argue that it is unnecessary to appeal to "selective sensitization" in the former case and "perceptual defense" in the latter, since recognition thresholds for

words have been shown to vary with their frequency of usage or familiarity (Howes and Solomon, 41) and this would provide a parsimonious explanation of both sets of findings. In the case of Allport-Vernon value systems, it seems reasonable to suppose that people with high theoretical values (and hence presumably with more courses, books, etc., in scientific fields) will have had more frequent visual contact with words like "logical" and "research"; in the case of taboo words, it also seems reasonable to suppose that the frequency of visual contact with words like "whore," "penis," and "bitch" is much less than with such control words as "child," "clear," and "dance." McGinnies' argument that these taboo words are much more frequent in ordinary conversations seems to be largely beside the point, since the test conditions were visual. The other main point raised by Howes and Solomon is that the emotional reactions *accompanied* recognition of the taboo words, appearing to precede simply because subjects inhibited reporting them—particularly since a member of the opposite sex was always present. These critics draw a delightful picture of what might have been going on in the subjects' minds during the McGinnies experiment. In an answer to these criticisms, McGinnies (66) draws a different picture. This is another issue yet to be resolved by further research. In this connection, a recent study by Lazarus and McCleary (61) reports that subjects may show heightened GSR to nonsense syllables previously associated with shock even when exposures are stopped prior to actual recognition. But here again one questions the validity of the demonstration—how can one react emotionally to the meaning of a sign before its significance has been appreciated? It may be that we will be forced to accept some conception of "unconscious" and "conscious" levels of perception or meaning.

Skinner (95) devised a "verbal summator" technique for studying language behavior which resembles these perception methods. Samples of meaningless speech sounds, obtained by permuting and combining elemental phonemes—a sort of verbal inkblot—are repeated until the subject himself perceives some meaningful form. According to Skinner, the verbal summator "evokes latent verbal responses through summation with imitative responses to skeletal samples of speech." That this method gets at the comparative strengths of verbal habits is indicated by the fact that the same double-log function of frequency to rank (Zipf's Law) appears when a large sample of such responses are analyzed. Estes (31) has described a visual form of summator which presents skeletonized verbal materials tachistoscopically.

#### *Association Methods*

Freud would have been the first to point out that the associations produced when a patient "allows one idea to lead to another" are in no sense "free" or random, but rather are semantically determined. An-

other analyst, Jung (48) used a more formal association method to get at the meanings of words to individuals. Lists of verbal stimuli calculated to touch off complexes were imbedded among neutral words. Among the indices of "unusual" responses was the rareness of the association itself. In order to judge the commonness or rareness of particular associations, it was necessary to know the comparative frequencies with which various responses to a given stimulus word occur in a representative sample of the population. Kent and Rosanoff (56) obtained responses to 100 common English nouns and adjectives from 1,000 subjects; their sample must have been fairly representative because the occasional re-checks that have been made show rather surprising agreement. Given norms like these, the unusualness of a subject's responses can be indexed by the frequency of occurrence of that response in the populations; *sharp* can be expected as a response to NEEDLE 152 times per 1,000 (15.2%), but *weapon* occurs only once (0.1%).

The gross majority of word associations are semantically determined, i.e., result from the mediation process set in motion by the verbal stimulus as a sign. All such associates are similar in some way to the stimulus word, either similar in meaning (NEEDLE-*pin*), which would include hierarchial relations (NEEDLE-*tool*), or in terms of commonness of context (NEEDLE-*thread*). The venerable associationistic principles of similarity and contiguity will be recognized here. After an intensive analysis of ways of classifying associates, Karwoski and Berthold (50) conclude that nearly all responses can be categorized as either some form of similarity or contrast. What about contrast responses? The single most frequent associate is often the direct opposite (LIGHT-*dark*; MAN-*woman*). For a number of reasons, this writer believes such contrast responses are *not* semantically determined at all, but rather reflect overlearning of verbal skill sequences, quite akin to FOOT-*ball*, APPLE-*cart*, and WASTE-*basket*. The tendency to free associate opposites increases with age, children readily giving similar and contextual responses but rarely opposites (cf., Woodworth, 114, p. 346). Furthermore, rather than being distributed among many varied but roughly equivalent words, as is the case with similar associates, the opposition tendency is largely restricted to a single word, the direct opposite (cf., Kent and Rosanoff tables). Karwoski and Schachter (54) report this same effect with opposites and add the fact that opposites are given with significantly shorter reaction-times than similars.

One of the more interesting applications of the association method has been in differentiating responses to sign, symbol, and object levels of stimulation. In an early study on this, Dorcus (29) compared associations to color words (signs) and actual bits of colored paper (objects).



Whereas co-ordinate and contrast responses were most common to color signs (WHITE-black, RED-blue), the names of contextually related objects were most commonly given to color objects (BLUE PAPER-ribbons; RED PAPER-fingernails). More recently Karwoski, Gramlich, and Arnott (51) have obtained associations to visually perceived actual objects, pictures of these objects, and verbal labels for these objects. The stimulus materials were such everyday things as *pipe*, *leaf*, *dollar*, and *pistol*. Where differences appeared, the dividing line was typically between the verbal level and the other two modes. For example, the most common response to the word FORK was, of course, *knife*—on both picture and objects levels the most common response was *eat*. Reaction times for the verbal level were also shorter.

Related to associational procedures are the effects of *context* upon meaning. It is a matter of common observation that a man's moods, emotions, and motives influence the character of his verbalizations. Bousfield and Barry (9) and Bousfield (8) found that subjects' rated moods (on a scale from "feeling well as possible" to "feeling as badly as possible") correlated with their rates of production of pleasant vs. unpleasant associates. The relatively stable attitudes of an individual also exert a contextual effect upon associations. Foley and MacMillan (33) have shown that associates to 40 ambiguous words (like *binding*, *administer*, *discharge*) are clearly influenced by the occupational status of subjects, as law students, medical students, or nonprofessional students. Perhaps because of the obviousness of the matter, no research seems to have been done upon the effect of the external, situational context upon meaning. This context includes the facial expressions and gestures of speakers, the objects present, and the activities underway, and so on (cf., Malinowski's [68] enlightening discussion of this in relation to decoding the language of another culture). Many slips of the tongue completely escape notice, simply because the situational context "carries" the intended meaning.

Howes and Osgood<sup>4</sup> have given attention to the manner in which the meaning of a particular sign is affected by the pattern of verbal materials within which it is imbedded. A sequence of four spoken words made up each item, the first three serving as the context and the fourth, spoken with greater emphasis, serving as the actual stimulus for word-association by the subjects. One experiment was designed to get at the effect of varying the *density* of contextual items having a common semantic direction: Group A heard three contextual stimuli of very similar meaning (e.g., *sinister*, *devil*, *evil*-DARK), Group B had one

<sup>4</sup> Howes, D. H., & Osgood, C. E. Studies on the combination of associative probabilities in linguistic contexts (in preparation).



neutral word added (*eat, devil, evil*-DARK), Group C two neutral words (*eat, basic, evil*-DARK), and Group D, as control, had all neutral words in the context. When the frequencies of response-words related to the particular context (e.g., *thief, mystery, dead*, etc.) were plotted as a function of the density of influence in the context, number of influenced associates turned out to be a simple multiple of the number of relevant words in the context. In a second experiment, three influencing words of relatively independent meaning were used as "context," and the question was how *temporal proximity* of contextual stimuli affects the meaning of the eliciting sign. With an item like *feminine, strong, young*-MAN, the responses clearly relevant to each contextual stimulus could be isolated (*woman, girl* vs. *hard, work* vs. *boy, child*, for examples). Then the frequency of occurrence of such related responses was plotted as a function of the order of presentation of their contextual stimuli (e.g., frequencies of *woman* as a response when "feminine" is in third position, nearest MAN, second position, and first position, most remote from MAN). The results indicate that degree of influence of a contextual stimulus upon the meaning of a sign is a sharply negatively accelerated function of the temporal interval between them. In other words, the influence of one word upon another falls off rapidly as the amount of intervening material increases.

### Scaling Methods

Considering the number of traits, abilities, and attitudes that psychologists have attempted to measure by scaling methods, it is significant that there has been practically no attempt to measure meaning this way. Since many psychologists must have thought about the problem at one time or another, this probably reflects the general belief that meanings are too complicated or too unique, or both. The few timid steps that have been taken in this direction involved drastic limitations on the scope of measurement, being aimed at scaling one or two isolated dimensions of meaning rather than meaning-in-general.

One group interested in scaling meaning has been the researchers in human learning, who wanted to be able to select materials for their experiments which could be specified with respect to this variable. A number of studies have been reported on the meaningfulness or "association value" of nonsense syllables (Glaze, 35; Hull, 42; Witmer, 113). The typical method was to use nonsense syllables as stimuli for word-associations, "meaningfulness" being indexed by proportions of subjects who could find any associations. One could then select equated lists for learning experiments, equated on this one basis at least.

Also motivated to provide learning experimenters with standardized

materials, Haagen (37) scaled 400 pairs of common adjectives in terms of their synonymity, vividness, familiarity, and association value. The method used was to have 280 college undergraduates judge these words on defined scales: (a) *synonymity* of a given word was judged on a seven-point scale in terms of the degree to which it denoted the same actions, objects or conditions as a standard word; (b) *vividness*, also judged on a seven-point scale, was defined as the clarity of graphicness of the impressions which a given word aroused; (c) *familiarity* was judged on a five-point scale, defined as the degree to which the judge knew the meaning of the word; and (d) *association value*, judged on a seven-point scale, referred to the degree to which the given word and a standard were associated in thought (e.g., hungry-thirsty, big-large, would have high association value). Useful though synonymity and associative value may be for purposes of learning experiments, they do not offer anything in the way of a measure of meaning—these judgments were always relative to some particular standard word, varying from one set of test words to another. The familiarity measure has nothing to do with meaning, of course. The vividness scale, being applied to each word separately rather than comparatively, probably is tapping some generalizable dimension of meaning.

Mosier (79) made the most direct application of scaling methods to the study of meaning. College subjects rated some 296 adjectives on an 11-point scale in terms of their favorableness-unfavorableness, these adjectives being selected from Thorndike's word lists as words expressing some degree of general evaluation. Frequency distributions of the responses to each word were scaled according to the method of successive intervals and plotted on probability paper. Plots for approximately 200 of these words were linear, indicating normal distribution of the data, when treated in this manner. Most of the words showed the "precipice effect" at one side or the other of the midpoint of the scale, indicating a higher degree of agreement on the *direction* (favorable or unfavorable) of the evaluation than on the *intensity*. Mosier was able to demonstrate a reasonable ordering of evaluative words in terms of their mean locations (e.g., excellent, good, common, fair, poor, etc.), including such information as the fact that "better" is connotatively less favorable than "good" (grammarians to the contrary). The most significant point is that Mosier demonstrated the feasibility of scaling certain aspects of meaning.

#### *Summary on Existing Methods*

The purpose of the preceding review has been to see if there already exist adequate methods of measuring meaning. By "adequate" I mean

already meeting most of the criteria of satisfactory measuring instruments. What are these criteria? (a) *Objectivity*. The method should yield quantitative and verifiable (reproducible) data. (b) *Reliability*. It should yield the same values within acceptable margins of error, when the same conditions are duplicated. (c) *Validity*. The data obtained should be demonstrably covariant with those obtained with some other, independent index of meaning. (d) *Sensitivity*. The method should yield differentiations commensurate with the natural units of the material studied, i.e., should be able to reflect as fine distinctions in meaning as are typically made in communicating. (e) *Comparability*. The method should be applicable to a wide range of phenomena in the field, making possible comparisons among different individuals and groups, among different concepts, and so on. (d) *Utility*. It should yield information relevant to contemporary theoretical and practical issues in an efficient manner, i.e., it should not be so cumbersome and laborious as to prohibit collection of data at a reasonable rate. While this is not an exhaustive listing of criteria, it is sufficient for our purposes.

1. The *physiological measures* (including action potential, GSR, and salivary records) are of somewhat dubious validity, since there has been no demonstration of the necessity of these peripheral components, and they are not sensitive measures in that we are unable to interpret details of the records in our present ignorance. Their chief drawback, however, is cumbersome—the subject has to be "rigged up" in considerable gadgetry to make such measurements. For this reason, even should validity and sensitivity problems be met satisfactorily, it seems likely that physiological indices will be mainly useful as criteria against which to evaluate more practicable techniques.

2. *Learning measures* (including semantic generalization and transfer/interference methods) are also somewhat cumbersome procedurally, but their main drawback as general measures of meaning is their lack of comparability. Any measure of generalization or interference is made with respect to the original learning of some standard which necessarily varies from case to case. The chief usefulness of learning measures, therefore, lies in the test of specific hypotheses.

3. The chief drawback with *perception measures* (e.g., what is perceived in ambiguous stimulus forms, the recognition-times for tachistoscopically presented words) is that they are not valid measures of *meaning*. They get at the availability or comparative habit strengths of alternate meanings or ways of perceiving. The fact that a religious person perceives VESPERS with a shorter presentation time than a theoretically oriented person says nothing about *how* the meaning of this term differs for them; the fact that the religious person perceives VESPERS more quickly than THEORY says nothing about the difference in meaning of these two words to this individual. The same statements apply to Skinner's "verbal summator" technique.

4. The selection of responses in *association methods* is partly dependent upon the meaning of the stimulus items (and hence indexes meaning) and partly dependent upon habit strength factors. The chief drawback, as a general measure of meaning, is lack of comparability. The responses of two individuals to the same stimulus, or of the same individual to two stimulus words, are

essentially unique as bits of data. Comparability can be obtained with group data, but this limits the method.

5. *Scaling methods* can be viewed as forms of controlled association in which the nature of the association is specified by definition of the scales (favorable-unfavorable, vividness, etc.) but the direction and intensity of association is unspecified. By the very nature of the scaling method, the comparability criterion is usually satisfied (provided the subjects can be shown to agree upon the meaning of the scale and its divisions). As used by Mosier, however, the method can have only partial validity. This is because he tapped only one dimension of meaning, the admittedly important evaluative dimension, whereas we know that meanings vary multidimensionally.

### THE SEMANTIC DIFFERENTIAL

The method to be proposed here is a combination of associational and scaling procedures. It is an indirect method in the same sense that an intelligence test, while providing objective and useful information, does not directly measure this capacity. However, unlike the intelligence test which treats this ability *as if* it were distributed along a single continuum (e.g., IQ scores vary along a single scale), we accept at the outset that meanings vary in some unknown number of dimensions and frame our methodology accordingly.

#### *Research Origins of the Method*

This method had its origins in research on synesthesia, defined by Warren in his *Dictionary of Psychology* (109) as "a phenomenon characterizing the experiences of certain individuals, in which certain sensations belonging to one sense or mode attach to certain sensations of another group and appear regularly whenever a stimulus of the latter type occurs." This implies a sort of "neural short-circuiting" that is present in only a few freak individuals, and it is true that many of the classic case histories in this area gave credence to this view: a subject reported pressure sensations about his teeth and cheeks whenever cold spots on his arms were stimulated (Dallenbach, 27); a girl displayed a rigid system of relations between specific notes on the musical scale and specific color experiences, consistent when tested over a period of seven and one-half years (Langfeld, 60). But here, on the other hand, was a man who imagined the number "1" to be yellow, "2" to be blue, "3" to be red . . . and, of course, "8" to be black (anyone who has played pool will recognize the origin of this system); and here was a little girl who recalled her friends as having pink faces and her enemies as having purple faces. What modalities are crossed in these cases?

A more recent series of investigations by Karwoski, Odbert, and their associates related synesthesia to thinking and language in general (cf., also Wheeler and Cutsforth, 115). Rather than being a rare



phenomenon, Karwoski and Odbert (52) report that as many as 13 per cent of Dartmouth College students regularly indulged in color-music synesthesia, often as a means of enriching their enjoyment of music. These photistic visualizers varied among themselves as to the modes of translation employed and the vividness of their experiences, and their difference from the general population appeared to be one of degree rather than kind. Whereas fast, exciting music might be pictured by the synesthete as sharply etched, bright red forms, his less imaginative brethren would merely agree that terms like "red-hot," "bright," and "fiery," as verbal metaphors, adequately described the music; a slow and melancholic selection might be visualized as heavy, slow-moving "blobs" of sombre hue and described verbally as "heavy," "blue," and "dark." The relation of this phenomenon to ordinary verbal metaphor is evident: a happy man is said to feel "high," a sad man feels "low"; the pianist travels "up" and "down" the scale from treble to bass; souls travel "up" to the good place and "down" to the bad place; hope is "white" and despair is "black." The process of metaphor in language as well as in color-music synesthesia can be described as the parallel alignment of two or more dimensions of experience, defined verbally by pairs of polar opposites, with translations occurring between equivalent portions of the continua (Karwoski, Odbert, and Osgood, 53, pp. 212-221).

Interrelationships among color, mood, and musical experiences were studied more analytically by Odbert, Karwoski, and Eckerson (81). Subjects first listened to 10 short excerpts from classical scores and indicated their dominant moods by checking descriptive adjectives arranged in a mood circle (cf., Hevner, 39). Then, on a second hearing, they listed the colors appropriate to each score. Significant relations were shown; the color associations to musical scores followed the moods created. A portion of Delius' *On Hearing the First Cuckoo in Spring* was judged leisurely in mood and preponderantly green in color; a portion of Wagner's *Rienzi Overture* was judged exciting or vigorous in mood and preponderantly red in color. When another group of subjects was merely shown the mood adjectives (with no musical stimulation) and asked to select appropriate colors, even *more* consistent relations appeared, suggesting that the unique characteristics of the musical selections had, if anything, somewhat obscured the purely verbal or metaphorical relations between colors and moods. Almost identical findings have been reported by Ross (92) for relationships between the colors used in stage lighting and reported moods produced in the audience. Data are also available for the effects of color upon mood in mental institutions and in industrial plants.

Responses to complex selections of music such as used in the above studies are themselves too complex for analysis of specific relations between auditory-mood variables and color-form variables. In order to get closer to the mechanisms of translation, Karwoski, Odber, and Osgood (53) used simple melodic lines recorded by a single instrument (clarinet) as stimuli. In a first experiment the subjects were typical photistic visualizers and they drew their photisms with colored pencils after hearing each short selection in a darkened room. The simplest

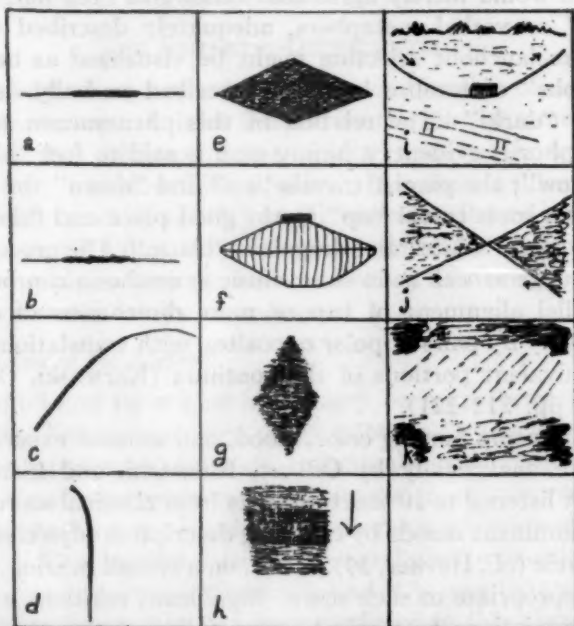


FIG. 2. SAMPLE OF PHOTISMS DRAWN BY COMPLEX SYNESTHETES TO REPRESENT A SIMPLE TONE WHICH GROWS LOUDER AND THEN SOFTER.

stimulus was a combination of crescendo and diminuendo on a single note—the sound merely grew louder, then softer—and this will serve to illustrate the results. As shown in Figure 2, subject *a* indicates increasing **LOUDNESS** by making the center of his line *heavier*, subject *b* by increasing *amplitude of vibration*, subjects *e*, *f*, and *g* by greater *thickness* of a solid form, subject *j* by more *concentrated focusing*, and subject *h* by more *saturated coloring* of the central portion. Subject *i* always created meaningful rather than abstract forms—here, a little car that comes *nearer* and then away again—yet the formal characteristics of his productions were generally like those of abstract synesthetes.

These are functionally or meaningfully equivalent responses to the same auditory stimulus dimension—i.e., there are alternate visual continua that can be paralleled with the loud-soft auditory continuum—and the advantage of the method is that its simplicity allows these relationships to show up clearly.

Are these photistic visualizers exercising a "rare" capacity or are they merely expressing overtly modes of translation that are implicit in the language of our culture? A second experiment used subjects who had never even thought of "seeing things" when they heard music (if they reported any such tendencies, they were eliminated). The same simple melodic lines as above were played and the subjects were instructed to "force themselves to draw something to represent what they heard." They produce the same types of visual forms and in approximately the same relative frequencies as the experienced visualizers. Finally, a group of 100 unselected students was given a purely verbal *meaning-polarity test*, each item of which appeared in the following form: LARGE-small; SOFT-LOUD, with instructions to circle that word in the second pair which "seems most clearly related to" the capitalized word in the first pair. Here again, essentially the same relations between music-mood variables and color-form variables discovered among sensitive synesthetes were linked meaningfully on the polarity test. *Large* was linked to *loud* by 96 per cent of these subjects, *near* with *fast* by 86 per cent, *bright* with *happy* by 96 per cent, *treble* with *up* by 98 per cent, and so on. It seems clear from these studies that the imagery found in synesthesia is on a continuum with metaphor, and that both represent *semantic* relations.

Are such semantic relations entirely dependent upon culture or is it possible that they reflect more fundamental determinants common to the human species? In an attempt to get at this question, the writer studied anthropological field reports on five quite widely separated primitive cultures—Aztec and Pueblo Indian, Australian Bushman, Siberian Aborigine, Negro (Uganda Protectorate), and Malayan—with the view of obtaining evidence on semantic parallelism. Special emphasis was given to nonmaterial aspects of culture (mythology, religion, arts, medical beliefs, birth, marriage, death complexes, etc.). The numerous pitfalls in the way of such analysis are probably obvious. Particularly, there is the danger of attributing relations to a primitive group when they are actually projections on the part of the observer or borrowings from the dominant Western culture. Therefore the results should be considered merely suggestive.

Nevertheless, the generality of certain relationships was quite striking. For example, *good* gods, places, social positions, etc., were

regularly *up* and *light* (*white*) in relation to *bad* things, which were *down* and *dark* (*black*). A prevalent myth tells of how the gods helped the original man to struggle "up" from the "dark," "cold," "wet," "sad" world below the ground to the "light," "warm," "dry," "happy" world on the surface of the earth. Among certain Siberian Aborigines, members of a privileged clan call themselves the "white" bones in contrast to all others who are referred to as "black" bones. And even among the Uganda Negroes we find some evidence for a white god at the apex of the hierarchy, and white cloth is clearly associated with purity, being used to ward off evil spirits and disease. Such data suggest the existence of a pervasive semantic frame of reference. Further study of the problem by more adequately trained investigators could be richly rewarding.

Stagner and Osgood (100) adapted this method and the logic underlying it to the study of social stereotypes. The notion of a continuum between the polar terms was made explicit by using such terms to define the ends of 7-step scales. Rather than studying the relations between continua, as above, a set of scales was used to measure the "meaning" of particular concepts, such as PACIFIST, RUSSIAN, DICTATOR, and NEUTRALITY. Successive samples of subjects were tested between April, 1940, and March, 1942 (including a sample obtained just prior to the Pearl Harbor incident). A single item on the tests appeared as follows:

PACIFIST: Kind : : : : : : : cruel

with the subject instructed to check that position on the scale which best represented the direction and intensity of his judgment. The concepts and scales related in successive items of the test were randomized to insure as much independence of judgment as possible. The feasibility and efficiency of using this method to record the changing structures of social stereotypes (e.g., the changing meanings of a set of social signs) were demonstrated. That a total shift from an essentially pacifistic to an essentially militaristic frame of reference had been accomplished, even before the Pearl Harbor incident provided the spark to overt expression, was clearly evident in the data.

More important from the point of view of methodology was the following observation: As used by our subjects in making their judgments, the various descriptive scales fell into highly intercorrelated clusters. Fair-unfair, high-low, kind-cruel, valuable-worthless, Christian-anti-Christian, and honest-dishonest were all found to correlate together .90 or better. This cluster represented, we assumed, a single, general factor in social judgments, the evaluative (good-bad) dimension of the



frame of reference. Gradients like strong-weak, realistic-unrealistic, and happy-sad were independent of this evaluative group and pointed to the existence of other dimensions within the semantic framework. Enforced shifts in the apparent reference point of the observer (by having subjects judge the same concepts "as a German" or "as an Englishman") produced gross and appropriate changes in the evaluative dimension but did not disrupt the qualitative pattern of each stereotype—e.g., the stereotype GERMANS, when judged by students playing the role of Germans, was still seen as relatively more "strong" and "happy" (remember, this was during 1940–1942) than "noble" or "kind." This illustrates the kind of difficulty experienced when one tries to assume the point of view of another (cf., Stagner and Osgood, 99).

### *Logic of the Proposed Method*

The researches described above gave rise to the following hypotheses:

1. *The process of description or judgment can be conceived as the allocation of a concept to an experiential continuum, definable by a pair of polar terms.* An underlying notion in our research is that these "experiential continua" will turn out to be reflections (in language) of the sensory differentiations made possible by the human nervous system. In other words, it is assumed that discriminations in meaning, which is itself a state of awareness, cannot be any finer or involve any more variables than are made possible by the sensory nervous system (cf., Boring, *The Dimensions of Consciousness*, 6). While failure to confirm this notion would not eliminate the proposed method as an index of meaning, its confirmation would greatly enhance the theoretical implications of this work.

2. *Many different experiential continua, or ways in which meanings vary, are essentially equivalent and hence may be represented by a single dimension.* This functional equivalence of many alternate continua was clearly evident in both the studies on synesthesia and those on the changing structure of social stereotypes. It is this fact about language and thinking that makes the development of a quantitative measuring instrument feasible. If the plethora of descriptive terms we utilize were in truth unique and independent of one another, as most philosophers of meaning seem to have assumed, then measurement would be impossible.

3. *A limited number of such continua can be used to define a semantic space within which the meaning of any concept can be specified.* From the viewpoint of experimental semantics, this both opens the possibility of measuring meaning-in-general objectively and specifies factor analysis as the basic methodology. If it can be demonstrated that a limited number of dimensions or factors are sufficient to differentiate among the meanings of randomly selected concepts, and if the technique devised satisfies the criteria of measurement stated earlier, then such a "semantic differential," as I have termed it, is an objective index of meaning. From the viewpoint of psychological theory, we may look upon the procedures followed in obtaining this measure as an operational definition of meaning, in the same sense that the procedures followed in obtaining the IQ score provide an operational definition of intelligence.

The operations followed in the present instance are explicit. They involve the subject's allocation of a concept within a standard system of descriptive dimensions by means of a series of independent associative judgments. The judgmental situation is designed to be maximally simple. Presented with a pair of descriptive polar terms (e.g., *rough-smooth*) and a concept (e.g., LADY), the subject merely indicates the direction of this association (e.g., LADY-*smooth*). We have developed two different methods for collecting data: In the *graphic method*, a pencil-and-paper technique which has the advantage that data can be collected from groups of subjects and hence very speedily, the subject indicates the intensity of his association by the extremeness of his checking on a 7-step scale. In the *judgment-time method*, which has the advantage that the subject cannot anticipate what concept is to be judged on a particular scale and hence cannot rationalize his reaction, intensity of association is indicated by the latency of the individual subject's choice reaction toward one or the other of the polar terms. In both methods each associative judgment of a particular concept against a particular descriptive scale constitutes one item. In successive items, concepts and dimensions are paired in deliberately rotated orders until every concept has been associated with every scale by every subject.

#### *A Factor Analysis of Meaning*

The procedures and results of this factor analysis will be described in detail elsewhere. A total of 50 descriptive scales, selected in terms of their frequency of usage, have been used in the judgment of 20 varied concepts, yielding a 1,000-item test. One hundred college students served as subjects. The graphic method was used.<sup>6</sup> The purpose of this factor analysis is to isolate a limited number of general dimensions of meaning having a maximal differentiating power, to try to bring some order out of semantic chaos. The larger the proportion of total variance in meaning accounted for by these factors, the more satisfactory will be the measuring instrument finally set up. A preliminary estimation of factors in the 50×50 matrix (each scale correlated with every other scale) indicates the existence of several roughly independent dimensions. An "evaluative factor" accounts for by far the largest portion of the variance. There is also evidence for a "strength factor," an "activity factor," and several others not clearly defined in this rough approxima-

<sup>6</sup> Apparatus for obtaining latency measurements from individual subjects has been constructed and will be standardized upon the reduced set of descriptive scales we hope to derive from this preliminary factor analysis. While this apparatus has the advantage that materials are projected from a film-strip and responses (directions and latencies) are photographed by a single-frame camera—all automatically—it is still applicable only to a single subject at a time and hence is time-consuming.

tion. Given such factors, it will be possible to select those specific scales (e.g., good-bad, strong-weak, active-passive, smooth-rough, hot-cold, etc.) which best represent them.

We have done some exploratory work on the use of the semantic differential as a practical measuring device. The two sets of profiles in Figure 3 will serve to illustrate the method. Two groups of only 20

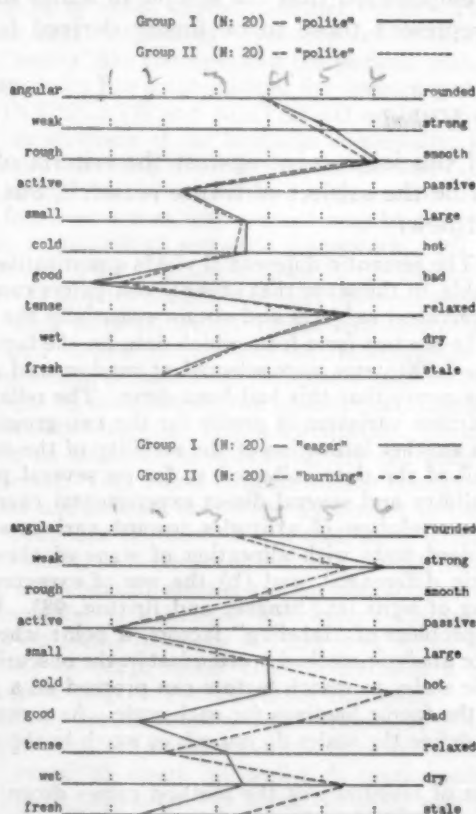


FIG. 3. ILLUSTRATION OF APPLICATION OF A PRELIMINARY FORM OF THE SEMANTIC DIFFERENTIAL FOR MEASURING THE CONNOTATIVE MEANINGS OF ADJECTIVES: A. UPPER PROFILES, MEDIANS FOR TWO GROUPS OF 20 SUBJECTS DIFFERENTIATING "POLITE"; B. LOWER PROFILES, MEDIANS FOR SAME TWO GROUPS OF 20 SUBJECTS DIFFERENTIATING "EAGER" AND "BURNING."

subjects each differentiated the meaning of the adjective "polite"; Group I also differentiated "eager" while Group II also differentiated "burning" (as part of a larger study). Median judgments of the 20 subjects on each scale are plotted. There is high agreement on the dif-

ferentiation of the same sign, "polite." The different, but somewhat similar signs, "eager" and "burning," show significant points of discrimination: whereas they are equally *strong* and *active*, "burning" is relatively *hot* and *dry* as compared with "eager," and "eager" is relatively *good* and *fresh* as compared with "burning." These differences are obviously what they would have to be if the method has any validity. It must be emphasized that the sample of scales shown here does *not* necessarily represent those to be finally derived from our factor analysis.

### *Evaluation of the Method*

Evaluation of this instrument against the criteria of measurement listed earlier will be the subject of future research, but some evidence can be presented now.

1. *Objectivity.* The semantic differential yields quantitative data which are presumably verifiable, in the sense that other investigators can apply the same sets of scales to equivalent subjects and obtain essentially the same result.

2. *Reliability.* In the test form from which data for the factor analysis were collected, 40 of the 1,000 items were selected at random and repeated. None of the subjects was aware that this had been done. The reliability coefficient was .85. The minimum variation in profile for the two groups of 20 subjects judging "polite" is another indication of the stability of the method.

3. *Validity.* All of the data collected so far on several problems display convincing face-validity and several direct experimental checks are planned. These include (a) correlation of attitudes toward various social objects as measured on standard tests with allocation of signs of these social objects within the semantic differential, and (b) the use of experimentally induced changes in meaning of signs (cf., Stagner and Britton, 98). We are not concerned about the problem of "labeling" factors, a point where the precision gained by the factor analytic method is often lost in the obscurities of language. Selection of specific scales to match factors can proceed on a purely objective basis, in terms of the factor loadings for each scale. As a matter of fact, the polar terms which define the scales do not admit much in the way of misinterpretation.

4. The question of *sensitivity* of the method comes down to whether it is able to reflect as fine distinctions in meaning as are ordinarily made. We have incidental evidence that a semantic differential can tease out nuances in meaning which are clearly felt but hard to verbalize deliberately.<sup>6</sup> If there is a real difference in the meaning of two signs, such that they would not be used in

<sup>6</sup> By way of illustration, most English-speaking Americans feel that there is a difference, somehow, between "good" and "nice" but find it difficult to explain. We gave several people these words to differentiate and it turned out that wherever "male" and "female" show a significant divergence, there also were "good" and "nice" differentiated (e.g., "good," like "male," is somewhat stronger, rougher, more angular, and larger than is "nice," which like "female" shifts toward the weak, smooth, rounded, and small directions of the space). Thus "nice man" has a slightly effeminate tone whereas "good woman" (as compared with "nice woman") has a narrowly moral tone.



precisely the same contexts, and if our measuring instrument includes a sufficient number of dimensions of the semantic space, then a significant difference should appear on at least one of the scales.

5. *Comparability.* It is here that the most serious questions arise. (a) *Is the method culture-bound?* If the tendency to dichotomize experiential continua is characteristic of Western culture but not necessarily elsewhere, then the method would not have generality. This is an empirical question requiring the skills of anthropologists and linguists for solution. (b) *Is the method limited to the differentiation of nouns against adjective scales?* The structure of our language is such that "adjectives" typically reflect abstracted qualities of experience and "nouns" the concepts and things dealt with. We have found it possible to set up scales like giant-midget, fire-iceberg, god-devil and to judge "concepts like INSINUATE and AGITATED against them. This does not seem "natural" to members of our language community, however; it is probably the stem or root meaning of words that our method taps. (c) *Can different concepts be compared?* To the extent that judgments of different concepts involve the same factor structure, any concept may be compared with any other against a single, standardized semantic framework. (d) *Can different individuals be compared?* This also comes down to the generality of the semantic factor structure. It is quite conceivable that different classes of people (scientists, ministers, etc.) have somewhat varied semantic structures, differing in the emphasis upon certain factors and interrelationships among them. In fact, a significant source of individual differences may lie here.

Our method can be criticized on the ground that it only gets at *connotative* meaning, not *denotative* meaning. This is a limitation. Both SIMON LEGREE and WAR might be allocated to approximately the same point in semantic space by our method. This would indicate similar connotative meaning, to be sure, but it would not indicate that these signs refer to the same object. Our differential will draw out the *hard, heavy, cold, ugly, threatening* connotations of the sign HAMMER, but it will not indicate that HAMMER is "an instrument for driving nails, beating metals, and the like, consisting of a head, usually of steel, fixed crosswise to a handle" (Webster's *Collegiate Dictionary*). In part, this limitation stems from our method of selecting descriptive scales in terms of frequency of usage rather than in terms of a logically exhaustive coverage, as given in Roget's *Thesaurus*, for example.

6. *Utility.* In any area of science, the development of an adequate method of measuring something (be it the wave length of radiation, blood chemistry, intelligence, or meaning) opens up well-nigh inexhaustible possibilities for application. (a) *Semantic norms.* In much the same way that Thorndike has established his norms for frequency-of-usage of common words in the English language, the semantic differential could be used to compile a functional lexicon of connotative meanings, a quantized thesaurus. Similarly, the gradual drift of changing meanings, both temporally and geographically, could be charted. (b) *Individual differences in meaning.* It is a truism that the meanings of socially significant signs differ for different classes of people. Concepts like CHURCH, LABOR LEADER, STALIN, and TRUMAN have different connotative significance to different people, and the semantic differential can be used to quantify these differences. In this sense, it can be used as a generalized, multidimensional attitude test. For example, 10 people may have identical degrees of favorableness toward NEGRO (evaluative dimension) and yet vary markedly with respect to other dimensions of the meaning-space. (c) *Changes*

in meaning. Under the pressure of events, the meanings of social signs change, e.g., the meaning of ITALIANS to Americans during the past half century. Similarly, under the "pressure" of psychotherapy, the meaning or emotional significance of certain critical concepts (e.g., FATHER, THERAPIST, ME, etc.) undergoes change. (d) *Quantification of subjective language data.* We have recently used the semantic differential as a means of scoring TAT reactions; not only is the testing process greatly speeded up, but the data are in easily manipulable form. Preliminary studies indicate that the essential individual differences in meaning of such projective materials, as teased out of complicated verbatim "stories," are sharply etched in the semantic differential data. (e) *Cross-cultural communication problems.* If the structure of the semantic space proves to be sufficiently general that the method can be translated into equivalent differentials in other languages, numerous possibilities are opened up. Are the fundamental factors in meaning and their relationships independent of the language spoken? Can the significant points of deviation in meaning of critical concepts, as between Americans and Russians, for example, be discovered? Can the finer, subtler degrees of acculturation into a new society be traced? And there are other potential applications, to aesthetics, to studying the development of meaning in children, and so on.

#### SUMMARY

The first portion of this paper describes a behavioral conception of the sign-process as developed from a general mediation theory of learning. The remainder is concerned with the problem of measuring meaning. Various existing approaches to the problem—physiological, learning, perception, association, and scaling methods—have been evaluated against the usual criteria of measurement and have been found inadequate. The development of a semantic differential as a general method of measuring meaning is described. It involves (a) the use of factor analysis to determine the number and nature of factors entering into semantic description and judgment, and (b) the selection of a set of specific scales corresponding to these factors which can be standardized as a measure of meaning. Using this differential, the meaning of a particular concept to a particular individual can be specified quantitatively as a particular point in the multidimensional space defined by the instrument. Some of the possible uses of such a measuring instrument are briefly indicated.

#### BIBLIOGRAPHY

1. ALLPORT, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937.
2. BARTLETT, F. C. *Remembering.* Cambridge, England: Cambridge Univer. Press, 1932.
3. BINGHAM, W. E., JR. A study of the relations which the galvanic skin response and sensory reference bear to judgments of the meaningfulness, significance, and importance of 72 words. *J. Psychol.*, 1943, 16, 21-34.
4. BLOOMFIELD, L. *Language.* New York: Holt, 1933.
5. BODER, D. P. The adjective-verb-

- quotient: a contribution to the psychology of language. *Psychol. Rec.*, 1940, 3, 310-343.
6. BORING, E. G. *The physical dimensions of consciousness*. New York: Appleton-Century, 1933.
  7. BOUSFIELD, W. A. An empirical study of the production of affectively toned items. *J. gen. Psychol.*, 1944, 30, 205-215.
  8. BOUSFIELD, W. A. The relationship between mood and the production of affectively toned associates. *J. gen. Psychol.*, 1950, 42, 67-85.
  9. BOUSFIELD, W. A., & BARRY, H., JR. Quantitative correlates of euphoria. *J. exp. Psychol.*, 1937, 21, 218-222.
  10. BOUSFIELD, W. A., & SEDGWICK, C. H. W. An analysis of sequences of restricted associative responses. *J. gen. Psychol.*, 1944, 30, 149-165.
  11. BOUSFIELD, W. A., & BARCLAY, W. D. The relationship between order and frequency of occurrence of restricted associative responses. *J. exp. Psychol.*, 1950, 40, 643-647.
  12. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, 42, 33-44.
  13. BRUNER, J. S., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, 27, 203-208.
  14. BRUNSWIK, E. Die Zugänglichkeit von Gegenständen für die Wahrnehmung. *Arch. ges. Psychol.*, 1933, 88, 377-418.
  15. CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visual perceived form. *J. exp. Psychol.*, 1932, 15, 73-86.
  16. CARROLL, J. B. Diversity of vocabulary and the harmonic series law of word-frequency distribution. *Psychol. Rec.*, 1938, 2, 379-386.
  17. CARROLL, J. B. The analysis of verbal behavior. *Psychol. Rev.*, 1944, 51, 102-119.
  18. CARTER, L. F., & SCHOOLER, K. Value, need and other factors in perception. *Psychol. Rev.*, 1949, 56, 200-207.
  19. CASON, H., & CASON, E. B. Association tendencies and learning ability. *J. exp. Psychol.*, 1925, 8, 167-189.
  20. CHOTLOS, J. W. Studies in language behavior: IV. A statistical and comparative analysis of individual written language samples. *Psychol. Monogr.*, 1944, 56, No. 2 (Whole No. 255), 75-111.
  21. CLARK, H. M. Conscious attitudes. *Amer. J. Psychol.*, 1911, 22, 214-249.
  22. CLARK, R. S. An experimental study of silent thinking. *Arch. Psychol.*, 1922, 7, No. 48. Pp. 101.
  23. COFER, C. N., & FOLEY, J. P., JR. Mediated generalization and the interpretation of verbal behavior. I. Prolegomena. *Psychol. Rev.*, 1942, 49, 513-540.
  24. COFER, C. N., JANIS, M. G., & ROWELL, M. M. Mediated generalization and the interpretation of verbal behavior. III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, 32, 266-269.
  25. COMSTOCK, C. On the relevancy of imagery to the processes of thought. *Amer. J. Psychol.*, 1921, 32, 196-230.
  26. CROSLAND, H. R. A qualitative analysis of the process of forgetting. *Psychol. Monogr.*, 1921, 29, No. 1 (Whole No. 130).
  27. DALLENBACH, K. M. Synaesthesia: "Pressury" cold. *Amer. J. Psychol.*, 1926, 37, 571-577.
  28. DARROW, C. W., & HEATH, L. L. Reaction tendencies relating to personality. In C. P. Stone, C. W. Darrow, C. Landis, and L. L. Heath. *Studies in the dynamics of*

- behavior. Chicago: Univer. Chicago Press, 1932.
29. DORCUS, R. M. Habitual word associations to colors as a possible factor in advertising. *J. appl. Psychol.*, 1932, **16**, 277-287.
  30. DUNCKER, K. On problem-solving. (Trans. by L. S. Lees from 1935 original.) *Psychol. Monogr.*, 1945 **58**, No. 5 (Whole No. 270).
  31. ESTES, W. K. A visual form of the verbal summator. *Psychol. Rec.*, 1940, **4**, 174-180.
  32. FOLEY, J. P., JR., & COFER, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 169-175.
  33. FOLEY, J. P., JR., & MACMILLAN, Z. L. Mediated generalization and the interpretation of verbal behavior: V. "Free association" as related to differences in professional training. *J. exp. Psychol.*, 1943, **33**, 299-310.
  34. FOLEY, J. P., JR., & MATHEWS, M. Mediated generalization and the interpretation of verbal behavior: IV. Experimental study of the development of interlinguistic synonym gradients. *J. exp. Psychol.*, 1943, **33**, 188-200.
  35. GLAZE, J. A. The association value of non-sense syllables. *J. genet. Psychol.*, 1928, **35**, 255-269.
  36. GOODWIN, J., LONG, L., & WELCH, L. Generalization in memory. *J. exp. Psychol.*, 1945, **35**, 71-75.
  37. HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453-463.
  38. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
  39. HEVNER, KATE. Experimental studies of the elements of expression in music. *Amer. J. Psychol.*, 1936, **48**, 246-268.
  40. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotional-ity and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229-234.
  41. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401-410.
  42. HULL, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730-734.
  43. HULL, C. L. *Principles of behavior: An introduction to behavior theory*. New York: Appleton-Century, 1943.
  44. JACOBSON, E. Electrophysiology of mental activities. *Amer. J. Psychol.*, 1932, **44**, 677-694.
  45. JOHNSON, W. Studies in language behavior: I. A program of research. *Psychol. Monogr.*, 1944, **56**, No. 2 (Whole No. 255), 1-15.
  46. JONES, H. E. Emotional factors in learning. *J. gen. Psychol.*, 1929, **2**, 263-272.
  47. JONES, H. E., & WECHSLER, D. Galvanometric technique in studies of association. *Amer. J. Psychol.*, 1928, **40**, 607-612.
  48. JUNG, C. G. *Studies in word-association*. (Trans. by M. D. Eder.) London: William Heinemann, 1918.
  49. KAPUSTNIK, O. P. The interrelation between direct conditioned stimuli and their verbal symbols (trans. from Russian title), 1930. *Psychol. Abstracts*, 1934, **8**, No. 153.
  50. KARWOSKI, T. F., & BERTHOLD, F., JR. Psychological studies in semantics: II. Reliability of free association tests. *J. soc. Psychol.*, 1945, **22**, 87-102.
  51. KARWOSKI, T. F., GRAMLICH, F. W., & ARNOTT, P. Psychological studies in semantics: I. Free association reactions to words, drawings, and



- objects. *J. soc. Psychol.*, 1944, 20, 233-247.
52. KARWOSKI, T. F., & ODBERT, H. S. Color-music. *Psychol. Monogr.*, 1938, 50, No. 2 (Whole No. 222).
53. KARWOSKI, T. F., ODBERT, H. S., & OSGOOD, C. E. Studies in synesthetic thinking: II. The roles of form in visual responses to music. *J. gen. Psychol.*, 1942, 26, 199-222.
54. KARWOSKI, T. F., & SCHACHTER, J. Psychological studies in semantics: III. Reaction times for similarity and difference. *J. soc. Psychol.*, 1948, 28, 103-120.
55. KELLER, MARGARET. Mediated generalization: the generalization of a conditioned galvanic skin response established to a pictured object. *Amer. J. Psychol.*, 1943, 56, 438-448.
56. KENT, GRACE H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, 67, 37-96, 317-390.
57. KOFFKA, K. *Principles of gestalt psychology*. New York: Harcourt Brace, 1935.
58. KÖHLER, W. *The mentality of apes*. (Trans. by E. Winter.) New York: Harcourt Brace, 1925.
59. KOTLIAREVSKY, L. I. Cardio-vascular conditioned reflexes to direct and to verbal stimuli (trans. from Russian title), 1936. *Psychol. Abstracts*, 1939, 13, No. 4046.
60. Langfeld, H. S. Note on a case of chromaesthesia. *Psychol. Bull.*, 1914, 11, 113-114.
61. Lazarus, R. S., & McCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, 58, 113-122.
62. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, 13, 283-293.
63. LYNCH, C. A. The memory values of certain alleged emotionally toned words. *J. exper. Psychol.*, 1932, 15, 298-315.
64. MCCLELLAND, D. C., & ATKINSON, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, 25, 205-222.
65. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, 56, 244-251.
66. MCGINNIES, E. Discussion of Howes' and Solomon's note on "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, 57, 235-240.
67. MAIER, N. R. F. Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, 35, 349-360.
68. MALINOWSKI, B. The problem of learning in primitive languages. Supplement in C. K. Ogden, & I. A. Richards. *The meaning of meaning*. New York: Harcourt Brace, 1938.
69. MARBE, K. *Experimentell-psychologische Untersuchungen über das Urteil*. Leipzig: Engelmann, 1901.
70. MASON, M. Changes in the galvanic skin response accompanying reports of changes in meaning during oral repetition. *J. gen. Psychol.*, 1941, 25, 353-401.
71. MAUSNER, B., & SIEGEL, A. The effect of variation in "value" on perceptual thresholds. *J. abnorm. soc. Psychol.*, 1950, 45, 760-763.
72. MAX, L. W. An experimental study of the motor theory of consciousness. III. Action-current responses in deaf-mutes during sleep, sensory stimulation, and dreams. *J. comp. Psychol.*, 1935, 19, 469-486.
73. MAX, L. W. An experimental study of the motor theory of consciousness. IV. Action-current responses in the deaf during awakening,

- kinaesthetic imagery, and abstract thinking. *J. comp. Psychol.*, 1937, 24, 301-344.
74. MELTON, A. W., & IRWIN, J. MCQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific response. *Amer. J. Psychol.*, 1940, 53, 173-203.
75. METZNER, C. A. The influence of preliminary stimulation upon human eyelid responses during conditioning and during subsequent heteromodal generalization. *Summ. Doct. Diss. Univ. Wis.*, 1942, 7, 152-154.
76. MILLER, G. A. *Language and communication*. New York: McGraw-Hill, 1951.
77. MORRIS, C. *Foundations of the theory of signs*. Internat. Encycl. Unif. Sci., Vol. 1, No. 2. Chicago: University of Chicago Press, 1938.
78. MORRIS, C. *Signs, language and behavior*. New York: Prentice-Hall, 1946.
79. MOSIER, C. I. A psychometric study of meaning. *J. soc. Psychol.*, 1941, 13, 123-140.
80. MOWRER, O. H. Preparatory set (expectancy): Some methods of measurement. *Psychol. Monogr.*, 1940, 52, No. 2 (Whole No. 233).
81. ODBERT, H. S., KARWOSKI, T. F., & ECKERSON, A. B. Studies in synesthetic thinking: I. Musical and verbal associations of color and mood. *J. gen. Psychol.*, 1942, 26, 153-173.
82. OGDEN, C. K., & RICHARDS, I. A. *The meaning of meaning*. London: Kegan Paul, 1923.
83. OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, 36, 277-301.
84. OSGOOD, C. E. An investigation into the causes of retroactive interference. *J. exp. Psychol.*, 1948, 38, 132-154.
85. POSTMAN, L., & BRUNER, J. S. Perception under stress. *Psychol. Rev.*, 1948, 55, 314-324.
86. POSTMAN, L., BRUNER, J. S., & McGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, 43, 142-154.
87. RAZRAN, G. H. S. Salivating and thinking in different languages. *J. Psychol.*, 1935-1936, 1, 145-151.
88. RAZRAN, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, 90, 89-90.
89. REED, H. B. The existence and function of inner speech in thought processes. *J. exp. Psychol.*, 1916, 1, 365-392.
90. RIESS, B. F. Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, 26, 238-240.
91. RIESS, B. F. Genetic changes in semantic conditioning. *J. exp. Psychol.*, 1946, 36, 143-152.
92. ROSS, R. T. Studies in the psychology of the theater. *Psychol. Rec.*, 1938, 2, 127-190.
93. SANFORD, R. N. The effects of abstinence from food upon imaginal processes: A preliminary experiment. *J. Psychol.*, 1936, 2, 129-136.
94. SANFORD, R. N. The effects of abstinence from food upon imaginal processes: A further experiment. *J. Psychol.*, 1937, 3, 145-159.
95. SKINNER, B. F. The verbal summator and a method for the study of latent speech. *J. Psychol.*, 1936, 2, 71-107.
96. SKINNER, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, 1, 69-76.
97. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century-Crofts, 1938.
98. STAGNER, R., & BRITTON, R. H., JR. The conditioning technique applied to a public opinion problem. *J. soc. Psychol.*, 1949, 29, 103-111.

99. STAGNER, R., & OSGOOD, C. E. An experimental analysis of a nationalistic frame of reference. *J. soc. Psychol.*, 1941, 14, 389-401.
100. STAGNER, R., & OSGOOD, C. E. Impact of war on a nationalistic frame of reference: I. Changes in general approval and qualitative patterning of certain stereotypes. *J. soc. Psychol.*, 1946, 24, 187-215.
101. SWITZER, S. C. A. Disinhibition of the conditioned galvanic skin response. *J. gen. Psychol.*, 1933, 9, 77-100.
102. THORNDIKE, E. L. *The teacher's word book*. New York: Teachers College, Columbia Univ., 1921.
103. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia Univ., 1944.
104. THORSON, A. M. The relation of tongue movements to internal speech. *J. exp. Psychol.*, 1925, 8, 1-32.
105. THOULESS, R. H. Phenomenal regression to the real object. II. *Brit. J. Psychol.*, 1931, 22, 1-30.
106. THUMB, A., & MARBE, K. *Experimentelle Untersuchungen über die psychologischen Grundlagen der sprachlichen Analogiebildung*. Leipzig: W. Engelmann, 1901.
107. TRAUGOTT, N. N. The interrelations of immediate and symbolic projections in the process of the formation of conditioned inhibition (trans. from Russian title), 1934. *Psychol. Abstracts*, 1935, 9, No. 1166.
108. TRAUGOTT, N. N., & FADEYEVA, V. K. The effect of difficult extinction of food-procuring conditioned reflexes upon the general and speech behavior of children (trans. from Russian title), 1934. *Psychol. Abstracts*, 1935, 9, No. 1167.
109. WARREN, H. C. (Ed.) *Dictionary of psychology*. New York: Houghton Mifflin, 1934.
110. WASHBURN, M. *Movement and mental imagery*. Boston: Houghton Mifflin, 1916.
111. WATSON, J. B. *Behavior, an introduction to comparative psychology*. New York: Holt, 1914.
112. WERTHEIMER, M. *Productive thinking*. New York: Harper, 1945.
113. WITMER, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, 47, 337-360.
114. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
115. WHEELER, R. H., & CUTSFORTH, T. D. Synaesthesia and meaning. *Amer. J. Psychol.*, 1922, 33, 361-384.
116. WYCZOIKOWSKI, A. Theoretical and experimental studies in the mechanism of speech. *Psychol. Rev.*, 1913, 20, 448-458.
117. ZIFF, G. K. *The psychobiology of language*. New York: Houghton Mifflin, 1935.
118. ZIFF, G. K. *Human behavior and the principle of least effort*. Cambridge, Mass.: Addison-Wesley, 1949.

Received August 29, 1951.

## THE VALIDITY OF THE BLACKY PICTURES<sup>1</sup>

GERALD S. BLUM

*University of Michigan*

AND

HOWARD F. HUNT

*University of Chicago*

Just as there is more than one way to skin a cat, there are many ways to try to validate behavior ascribed to a dog. Specific, definitive validity studies in the area of projective techniques are likely to continue to be infrequent by virtue of the fact that the instruments tend to be used in a variety of situations, each use requiring its own demonstration of validity. Conceivably a measuring device may prove quite effective in differentiating clinical syndromes and yet fall down in the assessment of dynamics as they operate within an individual. Or the theory underlying a given technique may make it more applicable to one type of problem than another. Offhand, these and similar complications may account in part for the current phase of confused ardor and recrimination concerning the value of our diagnostic tools.

But perhaps this very diversity may be turned into an asset through a kind of expedient, substitute solution to the criterion problem. While fully adequate criteria continue to be elusive, it seems reasonable to assume that a large variety of partially satisfactory criteria can serve temporarily as the best approximations to the "true" ones. In other words, a technique with merit should show up well when tested in a wide assortment of situations, despite the fact that no one situation provides a sufficient test by itself.

In the past the tendency has been to rely heavily on impressions of clinical utility as a basis for evaluating projective techniques. The Rorschach and Thematic Apperception Tests, for example, seem to have fared very well when judged in this way. While favorable opinions regarding clinical performance certainly deserve consideration, the dangers involved in judging validity solely on such a foundation are self-evident. Here, then, is another source of confusion which contributes to the current tempest over whether the Rorschach and other projective tests are "valid" or not.

Fortunately, the Blacky Pictures (3) are of such recent vintage that tradition and custom do not inevitably interfere with the development

<sup>1</sup> The authors wish to express their appreciation to E. Lowell Kelly for his helpful suggestions in the preparation of this manuscript.



of standards for evaluation. Unsystematic, subjective evidence based on clinical impressions need not play much of a role, if any, in the validation process. Of course, it is encouraging to learn that clinicians react positively to the test, just as it would be discouraging if the opposite tended to be the case. But acceptance of a technique as being "fruitful" hardly constitutes validation. An instrument must prove its worth in a variety of research settings.

The present article is intended to survey completed studies of the Blacky Pictures in a number of separate areas. These summaries, based on researches reported in detail elsewhere, have been assembled in order to provide some perspective on what has been done with the test to date and to suggest lines of future research. Sufficient evidence has accumulated to warrant such a stock-taking enterprise; and we shall take advantage of the opportunity to present our tentative appraisal of these available validity data.

#### THE BLACKY PICTURES TECHNIQUE

Before describing the various approaches relating to the validity of the Blacky Pictures, a short résumé of the technique itself is probably in order. The pictures were originally devised in 1947 as an attempt to get at the deeper recesses of personality in a more appropriate setting, geared directly to dynamic interpretation. It was felt that many projective techniques are lacking in theoretical orientation and consequently pose difficult interpretive problems. The twelve cartoons, tailored to fit psychoanalytic theory, portray the adventures of a dog named Blacky. The first introduces the cast of characters, which includes Blacky, Mama, Papa, and Tippy, a sibling figure of unspecified age and sex. Each of the subsequent cartoons is designed to depict either a stage of psychosexual development or a type of object relationship within that development. Dog cartoon figures were utilized to compensate for the increased structuring of the stimulus cards—the intention being to facilitate freedom of personal expression in situations where human figures might provoke an unduly inhibitory resistance as a result of being "too close to home." While on the one hand minimizing the dangers of resistance, the canine medium, thanks to Disney cartoons and comic strips, still preserves enough reality for subjects to identify themselves freely with the cartoon figures and to project their innermost feelings. It seems almost as if the animal cartoons appeal directly to the residues of childish, pre-logical thinking in adults, despite the added fact that they frequently realize consciously that they are telling about themselves.

The subject is asked to tell stories in response to the pictures. After his spontaneous story on each cartoon, he is asked a series of multiple-choice and short-answer questions pertaining to that psychoanalytic dimension. Upon finishing the inquiry on the last card, a third level of response is obtained by having the subject express his preferences for the various cartoons. For group administration the sexes are tested separately, since Blacky is presented as the "son" to males and as the "daughter" to females. An objective scoring system, recently revised on the basis of internal consistency of the four source scores (Spontaneous Story, Inquiry, Cartoon Preference, and Related Comments) for each dimension, is employed for research uses of the instrument. Subjects wind up with a profile of scores in terms of three possible degrees (0, +, ++ ) of disturbance on every dimension. When dichotomization of scores is desirable, the + and ++ categories are combined.

#### COMPARISON WITH THEORETICAL PREDICTIONS

The fact that the Blacky Pictures are based on psychoanalytic theory opens one kind of proving-ground for testing the instrument. The theory yields predictions which can be compared directly with results from the test. Completed studies of this type fall into three categories: syntactical relationships within the theory itself; application of the theory to demographic groups; and theoretical contributions to the understanding of various clinical syndromes. Research findings pertaining to the first two categories have already been published in detail (2) and will be summarized again only briefly. Four illustrations from the last category will then be described below.

##### *Syntax of the Psychoanalytic Theory of Psychosexual Development*

In the initial investigation, the Blacky Pictures were administered, in group form, to 119 male and 90 female students in elementary psychology classes at Stanford University. Analysis of the resulting protocols revealed statistically significant correlations (beyond chance expectancy) between dimensions on the test, which suggests that the dimensions are related to each other in a nonrandom manner. These significant correlations agreed with expectations based upon the internal syntax of the theory in those instances in which a definite prediction could be deduced from the theory.

##### *Application of the Theory to Demographic Groups*

As part of the same study, the Blacky protocols were also analyzed for sex differences. The number of obtained differences again clearly exceeded the total that would be expected to occur by chance alone.

These differences between the sexes were almost all in the direction predicted by psychoanalytic theory in those instances for which the theory permitted an unequivocal prediction.

### *The Theory as Applied to Various Clinical Syndromes*

*The paranoid schizophrenic.* Aronson (1), in a study aimed at the Freudian theory of paranoia, administered the Blacky Pictures individually along with several other psychological tests to 30 paranoid schizophrenics (Pa), 30 nonparanoid schizophrenic controls (Ps), and 30 normal controls (N). The three groups were equated for age, intelligence, occupation, religion, and veteran status. The two psychotic samples were differentiated according to the presence or absence of delusions, which was judged on the basis of all the available clinical materials. We shall summarize here only the section dealing with the Blacky Pictures. Significance tests computed from the blindly-scored, coded protocols revealed a large number of differences (including source scores and inquiry items as well as total dimensional scores) between the groups. The number expected to occur by chance alone could not be computed, since the various comparisons were not independent of each other (e.g., total dimensional score contained the separate sources). However, since the Blacky data were compared to predictions made from the theory, the issue of capitalization on chance was not deemed to be of major concern.

Within each test dimension (Oral Eroticism, Oral Sadism, etc.) attempts were made to integrate the findings, taking into account the level of awareness tapped by separate sources (Spontaneous Story, Inquiry, etc.). The following are excerpts from Aronson's summaries of the data (1, pp. 129-130):

1. *Oral Eroticism* (Prediction: Pa stronger than N). The paranoids showed more evidence of basic underlying oral deprivations than the normals. However, they tended, much more than the normals, to deny these deprivations. These results are in accord with the Freudian contention that paranoids are fixated at an early narcissistic stage of psychosexual development.

2. *Oral Sadism* (Prediction: Pa stronger than N or Ps). The paranoids exhibited very strong unconscious oral sadistic conflicts, as compared to the normals, but they did not differ from the psychotics. The paranoids also showed a great deal of difficulty in handling oral sadistic impulses at a near-conscious level. These results indirectly support Freud's statement that paranoids are fixated at a narcissistic stage of development. They directly support Klein and Rosenfeld's theory that oral sadistic fixations play an extremely important role in the development of paranoia.

3. *Anal Retentiveness* (Prediction: Pa stronger than N or Ps). The paranoids revealed much stronger anal retentive conflicts than the normals did. They also tended to exceed the psychotics in this respect. These results are in agreement with the prevalent analytic viewpoint that anal fixations are the primary ones in paranoia. The paranoids made little attempt to conceal their anal retentive

interests and attitudes. This frankness on their part, in contrast to their general evasiveness on the Blacky technique, was probably due to the fact that anal retentive character traits are rather highly valued in Western society, and hence do not have to be denied. Finally, according to analytic theory, anal fixations frequently lead to the adoption of passive (homo)sexual aims.

Summarizing the remaining dimensions, he found the paranoid group showing greater evidences of masturbation guilt; conscious attempts at denial of strong underlying castration anxiety; a consistent tendency toward feminine identification; severe superego conflicts; and a preference for narcissistic types of love-object choice. Little or no differences were obtained on the dimensions of Anal Expulsiveness, Sibling Rivalry, and Ego-Ideal. Some indirect evidence of stronger oedipal fixations in the paranoid group appeared on the Anacletic Love-Object dimension, but not on Oedipal Intensity.

*The stutterer.* Merchant (8) utilized the Blacky Pictures to explore the psychosexual development of stutterers. Blacky protocols (group administration) were obtained from 20 stutterers and 20 non-stutterers attending a middle-western college. Every stutterer was matched with a non-stuttering student of the same sex, age, family constellation, and classification on the Ohio College Aptitude Test. There were 15 men and 5 women in each group. She found that the stutterers differed significantly from the controls on 7 of the 14 Blacky dimensions. The experimental group showed more evidence of disturbance on Oral Eroticism, Oral Sadism, Castration Anxiety and Penis Envy, Identification Process, Guilt Feelings, and Anacletic Love-Object. On the dimension of Anal Expulsiveness the stutterers received significantly fewer strong scores. Merchant evaluates her findings as follows:

The Blacky Test presents a fairly clear-cut picture of the psychosexual development of these stutterers, and one in agreement with psychoanalytic theory of stuttering according to Fenichel. Developmental status of stutterers is sufficiently consistent, or the Blacky Pictures sufficiently sensitive, or both, that these characteristics did show up with only 20 pairs of subjects. . . .

The final paragraph of such a pilot study must be one of "Suggestions" rather than of "Conclusions," but it may suggest rather strongly that here is evidence for the psychoanalytic theory of stuttering and for the Blacky Pictures as an instrument useful in the investigation of psychosexual development.

*The peptic ulcer case.* In a study reported by Blum and Kaufman (4) the Blacky Pictures were administered, in group form, to 14 adult male ulcer patients as one phase of a total psychological workup. The results are summarized as follows:

Exploration of the scored responses, in conjunction with those of three control groups (normals, paranoid schizophrenics, and nonparanoid schizophrenics), uncovered two opposite trends within the ulcer sample. Whereas all the ulcer cases wrote strongly oral stories, only one-half selected multiple-choice



alternatives indicating oral conflict in the inquiry items on Cartoon I. The discrepancy between close-to-conscious expression of oral needs in the one half, contrasted with obvious attempts at denial in the other, suggested the hypothesis that there may be two very different patterns of ulcer dynamics. Further investigation of the "primary" and "reactive" subgroups showed the latter to be consistently more evasive in the content of their stories on the first oral cartoon. Comparisons on the remaining Blacky dimensions revealed the "primary" pattern to consist of a relative absence of conflicts centering about anal retentiveness, oedipal feelings, castration anxiety, and narcissistic object-choice, along with an excessive concern for mothering and little conscious desire for strong, masculine identification. The "reactive" pattern, closer to the popular conception, is characterized by repressed receptive longings, intense, unresolved oedipal feelings, the conscious wish to emulate a decisive father figure, a pervasive sense of guilt, and a highly narcissistic approach to others.

The formulation of these two patterns, based upon a limited number of cases, was presented in the form of a suggestion for the direction of additional research in the area of peptic ulcer dynamics. Differences between the two subgroups on other psychological tests are reported in a paper by Marquis, Sinnett, and Winter (7).

*The sexual offender.* Lindner<sup>2</sup> reports what he describes as a "validation study" of the Blacky Pictures done in two Maryland State prisons. The technique was administered individually to 67 male sexual offenders and 67 nonsexual offenders matched according to age, race, I.Q., education, marital status, socioeconomic status, length of sentence, and previous commitments. The experimental group was comprised, in decreasing order of frequency, of cases of homosexuality, sodomy, rape, pedophilia, exhibitionism, carnal knowledge, and contributing to the delinquency of minors. Analysis of the data revealed the sexual offenders to be significantly more disturbed than the nonsexual offenders on 9 of the 13 Blacky dimensions. These groups were differentiated at the .001 level of significance on Oral Eroticism, Oedipal Intensity, Masturbation Guilt, Castration Anxiety, Sibling Rivalry, and Guilt Feelings; at the .01 level on Oral Sadism and Narcissistic Love-Object; and at the .05 level on Anaclitic Love-Object. Anal Expulsiveness also approached significance ( $P < .10$ ), whereas Anal Retentiveness, Positive Identification, and Positive Ego Ideal did not discriminate between the sexually deviant group and the nonsexually deviant controls. Within the experimental sample of offenders there were no significant differences according to classification.

Lindner presents the following interpretation of his findings:

In conclusion, then, we have found that the Blacky Pictures test is a valid indicator of psychosexual deviation in a selected population. It is sufficiently

<sup>2</sup> LINDNER, H. A study of imprisoned sexual and nonsexual offenders by means of the Blacky Pictures test. (Presented at the 1951 annual meeting of the Eastern Psychological Association.)

*this is longitudinal*

sensitive to discriminate between two groups of subjects: a sexually aberrant group and a nonsexually aberrant group. To the extent that this test represents psychoanalytic theory, these data may be considered to support such theory as a plausible rationale.

#### VALIDATION BY EXPERIMENTAL TECHNIQUES

The preceding section summarized a series of studies in which Blacky data were compared with predictions derived from psychoanalytic theory. A second type of approach to the validity of the instrument lies in comparisons with independent experimental measures. This method is illustrated by the case of a nine-year-old girl who served as a subject in a broad research program aimed at the psychoanalytic theory of the "oral character" (5, 6). As one phase of the research, the Blacky Pictures were administered to all 26 children in the fourth grade of an elementary school in Ann Arbor, Michigan. In the process of checking the adequacy of test administration a Blacky record (all identifying marks removed) was randomly selected for a supervisory consultation. The stories told by this girl revealed very strong involvement on Cartoon I (Oral Eroticism) and spontaneous oral references on seven of the remaining ten pictures. The striking evidence of preoccupation with orality suggested a handy way to validate her test performance, since data for the class were available on a large number of experimental measures of oral character structure.<sup>3</sup> These measures had been obtained from experiments, time-sampling, teacher ratings, and socio-metrics.

Out of a total of 20 variables our heroine fell in the most oral third of the group on 16 variables. On 12 of the 20 variables she ranked between first and fifth. For example, she was fifth in the class of 26 on average number of one-ounce cups of ice cream consumed daily; first on teacher ratings of "needs praise from others"; third in taste suggestibility; third in the inability to say "No" to requests from others; and so on. It is also interesting to note that on two nonoral control variables—neatness and collecting tops of cups during the ice cream experiment—she ranked 23rd and 13th respectively.

The fact that independent experimental evidence clearly substantiated the blind Blacky prediction of strong oral involvement in this girl should encourage such approaches to validation.

<sup>3</sup> BLUM, G. S. Validation of a child's Blacky record by experimental techniques. (Presented at the round table on the Blacky Pictures held at the 1950 annual meeting of the American Psychological Association.)

## PREDICTION OF BEHAVIOR IN A GROUP SETTING

Swanson (10), in an attempt to analyze the relationship between personality characteristics of group members and their observed behavior in the group, administered the Blacky Pictures to each of two groups of 20 delegates at the 1949 session of the National Training Laboratory for Group Development in Bethel, Maine. These groups contained roughly equal numbers of men and women, with a median age of 47. Independent data on the behavior of participants and groups were secured as part of the general routine of the project and were not obtained specifically for this study.

The Blacky protocols were later scored in a blind analysis by the author of the test, who had available only the age and sex of the subject. A profile of scores along the various psychoanalytic dimensions was thus available for each subject. The three conventional categories of very strong, fairly strong, and weak or absent were employed, so that one individual might have gotten a ++ on Oral Eroticism, a + on Oral Sadism, a 0 on Anal Expulsiveness, and so on down the line. Training Groups A and B were both scored in this fashion.

Swanson, from his knowledge of psychoanalytic theory, made predictions concerning the pattern of Blacky dimensional scores which should be associated with various tendencies. For example, he reasoned that strong scores on Oral Sadism, Anal Expulsiveness, Oedipal Intensity, Sibling Rivalry, and Guilt Feelings should raise the total amount of an individual's actual participation in a permissive group like A, whereas Oral Eroticism, Anal Retentiveness, and Anacletic Love-Object choice should lower it. Using this intuitive pattern analysis of Blacky scores, he ranked the members of the group on predicted participation and then correlated the rankings with observed actual participation. The patterns of predictions based on Blacky scores for Group B had to be revised somewhat in view of its markedly different atmosphere. In contrast to Group A, which was extremely permissive and productive in its decision-making, B had a climate of hostility, which was accompanied by lower participation and less effective decision-making.

The main results of the study are presented in Table 1, which is reproduced from Swanson's report. In commenting on the utility of the Blacky Pictures for his design, he writes:

Unsystematic though our data are, they suggest that the Blacky Test is a suitable instrument for recording the kind of personality variables that are important in interaction. Neither the number of our groups, nor their size, nor the

variety of predictions attempted, permit anything like a final verdict on its adequacy, but it seems to merit further use when such an instrument is needed in research designs.

What has been especially impressive in working with these personality variables is the utility of an interpersonal interpretation of psychoanalytic theory. Along with its technical ingenuity, much of the merit of the Blacky Test seems

TABLE 1  
SUMMARY OF RELATIONSHIPS BETWEEN THE BLACKY PREDICTIVE PATTERNS  
AND THE CRITERION VARIABLES FOR GROUPS A AND B\*  
After Swanson (10)

Criterion Variable	Clinical Prediction Based on Blacky Scores	
	Group A	Group B
Total volume of participation	.63§	.56†
Percentage of total participation expressed that is disagreement and modification	.21	.32
Percentage of total participation received that is disagreement and modification	.33	.57†
Percentage of total liking choices that are favorable	.68§	.49†
Percentage of total mentions of being influenced that are positive in direction	.36	.75§
Percentage of total mentions of being a source of in- fluence that are positive in direction	.52†	.55†
Ideology Test—total score	.56†	.33
Desire to be a leading member of the group	.68§	.37
Desire to conform to the group's wishes	.50†	no data
Satisfaction with own performance in the group	††	no data

\* Except in the one case noted, all measures in this table are rhos. *N* is 17 in each group.

† Relationship measured by the use of simultaneous probability.

‡ Significant at the .05 level.

§ Significant at the .01 level.

to rest on the fact that, unlike most other personality depth instruments, with such exceptions as the Rorschach, it originated in a systematic theory that tries to handle the individual mechanisms of interpersonal relations in a single, comprehensive scheme. Unquestionably that scheme is not infallible, but it is systematic and testable.

#### THE CLINICIAN'S JUDGMENT AS A STANDARD OF COMPARISON

The clinical approach, in which the clinician himself fulfills the role of criterion agent, is a more conventional procedure in general use. To



the authors' knowledge, no systematic Blacky explorations of this sort have already been completed. However, we may be able to set our sights more sharply by considering three "exercises" which fall into this category.

*The therapist's judgment as the standard.* One such investigation has been discussed in the preface to the sample case workup in the Blacky Pictures Manual.<sup>4</sup> The procedure, in which a patient's therapist was asked after 10 hours of therapy to pick out his patient from among a group of five, is described as follows:

All of the patients had been tested and the analyses written up blindly before therapy was begun. The therapist made choices first from five short descriptions (laid out in random order) adapted from each dimension, and finally from the five complete personality evaluations. On all of the specific dimensions he succeeded in identifying his patient by either first or second choice, and in the final analysis he made the correct first choice (3, p. 8).

A similar study along more extensive lines is now being conducted by Ellis in which the Blacky Pictures have been administered to a patient with some two hundred hours of psychoanalysis.<sup>5</sup> Ratings by therapist and patient, without knowledge of Blacky results, are being compared with ratings made by 30 clinical psychologists from the Blacky protocol alone.

*The clinician's prediction of the patient's test performance.* A variation in the use of the clinician as a criterion agent is to ask him to predict, on the basis of all available information, how the patient will perform on the test. An illustrative exercise was done at VAH Fort Custer in which an extensive set of interview and test materials (Rorschach, TAT, Bender-Gestalt, Wechsler-Bellevue) was used to predict presence (+) or absence (0) of disturbance on each of the Blacky dimensions.<sup>6</sup> The predicted and actual scores for this case coincided in 12 out of 13 instances insofar as gross indication of presence or absence of disturbance was concerned.

*The clinical case study.* Michal-Smith, Hammer, and Spitz (9) report the case study of a nine-year-old Negro boy referred to a clinic for stealing toys. Case material revealed a history of spying on his mother and stepfather at night, masturbation, homosexual and heterosexual episodes. Excerpts from the patient's Blacky stories dramatically confirm and elaborate the severe oedipal ambivalence expected on the basis of the history.

<sup>4</sup> The clinician whose cooperation made this exercise possible was Dr. Abraham Carp, Chief Psychologist, VAH, Fort Custer, Michigan.

<sup>5</sup> Personal communication.

<sup>6</sup> The clinician in this case was Justin L. Weiss, VA trainee in clinical psychology, University of Michigan.

## SOME CRITICAL COMMENTS

At first glance, the research cited above might be interpreted as pretty clearly establishing the validity of the Blacky Pictures. For various reasons, however, these results should be considered as encouraging and suggestive rather than definitive. They encourage further exploration of both the test and the personality theory it serves; they strongly suggest that "there is something there," but do not necessarily indicate "what it is," or "where it is." Two examples from the original monograph illustrate methodological difficulties which constrain a definitive interpretation:

1. The only theoretical predictions which could be compared with empirical findings were those predictions which, on the basis of their content, appeared to be related to the differences or correlations obtained on the test. How many predictions could the theory generate? And what proportion of this hypothetical population of predictions was actually checked and found to agree with data from the test? Nobody can tell; psychoanalysis is still insufficiently logical and systematic to permit an exhaustive determination of all possible, meaningful predictions. The theory was carefully combed and those predictions worked out which seemed pertinent to the dimensions of the test. Some, of course, might have been missed. In addition, only statistically significant differences and correlations in the data from the test were compared with predictions from the theory; nonsignificant differences and correlations, though they might have been quite legitimately reflecting absence of relationship or difference, had to be left out of the comparisons because of their ambiguity (2, p. 23).

Thus we see that only certain kinds of data from the test entered the comparisons with what was probably only an incomplete population of theoretical predictions. The total possible number of comparisons between theory and test is unknown, so a rigorous determination of what constitutes chance agreement between theory and test cannot be made. In these circumstances, one cannot determine, precisely, whether obtained agreement between theory and test significantly exceeds the "real" chance expectancy for agreement, though there is no question that it exceeds the most reasonable estimate of expectancy that we can make at the moment.

2. For another example, let us turn to the test. How should one interpret the obtained statistically significant sex differences and dimensional inter-correlations? Is each significant difference "real" because it is statistically significant? Are all of them "real" just because more than a chance number of them seemed to occur? Not necessarily. The statistical analysis in this first study could do no more than indicate an apparent excess of such significant differences; it could not indicate which were "real" and which attributable to chance. For this reason, all must be considered to have only a tentative sort of status until cross-validating studies, on other samples, reveal which differences and correlations survive the vicissitudes of sampling and random error, and which do not. In effect, the results must be interpreted as hypotheses to be tested in later studies under a variety of conditions.

One may find the seeds of ad hoc reasoning, capitalization on chance, and kindred sins in the subsequent researches summarized above. These methodological frailties frequently, often unavoidably, characterize exploratory research in this area—hence the emphasis here on the suggestive and encouraging, but not definitive, character of the current findings on the validity of the Blacky Pictures. It is noteworthy, however, that these studies investigated the test performance of various kinds of persons, ranging from psychiatric cases to normal individuals, and employed criteria as diverse as psychiatric symptomatology, behavior in a group, and amount of ice cream eaten. In a very important sense, these studies constitute a species of cross-validation of the test; and, fortunately, they all tend to yield positive and confirmatory results. As such evidence accumulates, one's confidence in the test can justifiably increase sharply. It becomes a bit difficult to consider that sampling errors, capitalization on chance, and the like, by themselves, are responsible for the positive results obtained under such varied circumstances.

Of course, the full range of possibilities for validation research has by no means been explored as yet. Some studies should employ, more directly, the type of data upon which psychoanalytic theory is based—the events of analytic therapy. Up to the present, no systematic investigation has compared performance on the pictures with such criteria as predictions and judgments made by trained analysts, the course and content of the patient's associations, the vicissitudes of the transference and so on. Since these are the data psychoanalysis was invented to order and comprehend, test and theory could then be compared at a somewhat less derivative and inferential level than has been possible heretofore. Many of the relevant and even critical aspects of personality may be detected less in the form of overt behavior on the part of the patient than through hunches, feelings, impressions, and even personal reactions on the part of the therapist. More refined research in this area demands the use of therapists and observers as sensitive discriminators. With appropriate study, these sources could yield criterion data relevant to those inchoate, unexpressed aspects of personality so dimly and ambiguously reflected in overt symptoms, social behavior, and the like.

The close articulation of the Blacky Pictures with psychoanalytic theory may or may not prove to be a great virtue. Both are in their scientific infancy; both really are still in the "invention" stage of development. If the theory, or something like it, turns out to be essentially

correct, the advantages of having a device like the Blacky are obvious. If the theory proves to be grossly wrong in either metaphor or syntax, Blacky's prognosis also will be bleak, unless he succeeds in demonstrating his discriminatory power in experimental, behavioral studies of personality which do not depend entirely upon psychoanalytic theory.

At present, one cannot be sure of the ultimate fate of psychoanalysis as a coherent, logical, and systematic theory. Nor can one forecast Blacky's future with certainty. The instrument needs much additional research and exploration. Its early publication for general use was designed to facilitate such activity and to broaden its base. In the present article we have sought to convey recent findings and to achieve some perspective in their evaluation.

## REFERENCES

1. ARONSON, M. L. A study of the Freudian theory of paranoia by means of a group of psychological tests. Unpublished doctoral dissertation, University of Michigan, 1950.
2. BLUM, G. S. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, 3, 3-99.
3. BLUM, G. S. *The Blacky Pictures: A technique for the exploration of personality dynamics*. New York: Psychological Corp., 1950.
4. BLUM, G. S., & KAUFMAN, JEWEL B. Two patterns of personality dynamics in male ulcer patients, as suggested by responses to the Blacky Pictures. *J. clin. Psychol.*, in press.
5. BLUM, G. S., & MILLER, D. R. Exploring the psychoanalytic theory of the "oral character." Presented at the 1950 APA symposium entitled "Experimental Approaches to Psychoanalytic Theory." *J. Pers.*, in press.
6. BLUM, G. S., & MILLER, D. R. Exploring the psychoanalytic theory of the "oral character": a follow-up study. *Amer. Psychologist*, 1951, 6, 339-343. (Abstract)
7. MARQUIS, D. P., SINNETT, E. R., & WINTER, W. D. A psychological study of peptic ulcer patients. *J. clin. Psychol.*, in press.
8. MERCHANT, FRANCES C. Psychosexual development in stutterers. *J. consult. Psychol.*, in press.
9. MICHAL-SMITH, H., HAMMER, E., & SPITZ, H. Use of the Blacky Pictures with a child whose oedipal desires are close to consciousness. *J. clin. Psychol.*, 1951, 7, 280-282.
10. SWANSON, G. E. Some effects of member object-relationships on small groups. *Human Relations*, in press.

Received August 25, 1951



## A MEASURE OF RELATION DETERMINED BY BOTH MEAN DIFFERENCE AND PROFILE INFORMATION

CHARLES E. OSGOOD AND GEORGE J. SUCI

*University of Illinois*

In psychology and other social sciences the investigator often has available a number of different measurements, constituting a profile, for each of several variables. The relationship between any two such variables is usually expressed by a product-moment correlation or some derivative of this ratio. This expression of relationship reflects only the *profile* similarity between the two sets of measures; information regarding the absolute *differences*, or discrepancies, between the means of the measurements is discarded. In many cases this difference information is irrelevant—when, for example, one correlates height (measured in inches) with weight (measured in pounds), or when one correlates two aptitude tests in which the scores depend quite arbitrarily upon the number of items in each battery. But there are many other cases where this difference information is relevant. Measurement specialists have often expressed concern, for example, over the fact that a moron and a genius may have closely correlated profiles, despite the absolute discrepancy between their scores. During the past three years the writers have been working toward the development of objective methods of measuring connotative meaning, using systems of descriptive scales as the bases for judgments.<sup>1</sup> As will be shown in subsequent examples, usual correlation procedures failed to give valid representation of semantic relations, and we were forced to search for other measures of relationship. The *difference method* we have developed, while particularly applicable to semantic measurements, is by no means restricted to this type of data. Problems identical with ours exist in many areas of social science,<sup>2</sup> and probably elsewhere as well.

<sup>1</sup> OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, 49, 197-237. A preliminary description of the difference method reported here was given in two papers at the meeting of the American Psychological Association, September, 1950.

<sup>2</sup> As evidence of concern over such problems we may cite the following: R. B. Cattell ( $r_p$  and other coefficients of pattern similarity. *Psychometrika*, 1949, 14, 279-298) discusses exactly the same issue in connection with the measurement of differences in culture patterns, deriving a "coefficient of pattern similarity" which is similar to the method described here; Cronbach has recently described (mimeographed materials) a measure of relation which appears to be identical with that described here, but developed in connection with interpersonal measurements from sociometric and Q-technique studies.

# THE DIFFERENCE METHOD AND THE GEOMETRIC MODEL ON WHICH IT IS BASED

Before describing the difference method it will be useful to clarify the nature of the problem. Figure 1 shows five seven-point scales. In our own work on semantic measurement such scales are defined by polar adjectives. When subjects are instructed to rate each of a number

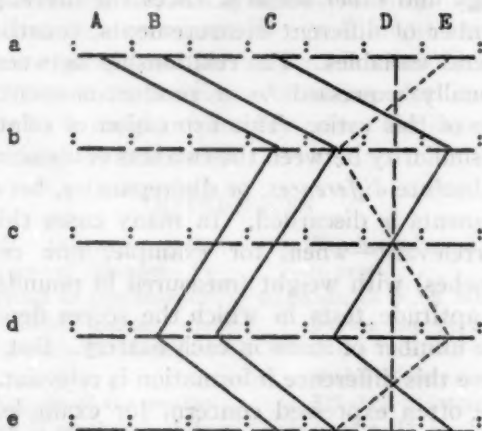


FIG. 1. PROFILES FOR FIVE HYPOTHETICAL VARIABLES (A, B, C, ETC.) JUDGED ON FIVE SCALES (a, b, c, ETC.).

VARIABLES A, B, AND C CONSTRUCTED SO AS TO HAVE IDENTICAL PROFILES; VARIABLES C, D, AND E CONSTRUCTED SO AS TO HAVE NEARLY IDENTICAL MEAN DIFFERENCES FROM A.

of concepts on each of the scales (e.g., *LADY rough-smooth*, *SIN high-low*, *FATHER strong-weak*, etc.), profiles like those shown are generated. Concepts are represented by capital letters (A, B, C, etc.) and scales for judgment by lower-case letters (a, b, c, etc.). Suppose A represents LOVE, B represents AFFECTION and C represents HATE. Since they covary perfectly, despite the gross absolute discrepancies, intercorrelations among them would all be 1.00, leading to the inference that LOVE is as similar in meaning to HATE as it is to AFFECTION. The correlation of any other concepts with D would be indeterminate since the variance of D is zero, yet concept D (e.g., AGGRESSION) is clearly closer to C (e.g., HATE) than to either A or B. Thus it can be seen that the product-moment correlation may not only distort the information but may be completely inapplicable in some cases. What is needed for these and similar kinds of data is a measure of relationship which takes into account both the mean discrepancies and the profile

covariation, thereby providing fuller reflection of the total information in the measurements.

If we have  $n$  variables and  $k$  measurements, every variable can be considered a point fixed in  $k$ -dimensional space by the  $k$  coordinates. Data of this kind can be arranged in matrix form as shown in Figure 2.

	$j$	$1$	$n$
$1$	$i j$	$i 1$	
$k$			

FIG. 2. GENERALIZED SCORE MATRIX FOR  $n$  VARIABLES MEASURED ON  $k$  DIMENSIONS.

Let us define the following symbols:

- $X$  = a measurement in cell  $ij$  in raw score form;
- $\bar{x}_j$  = a measurement in cell  $ij$  standardized with respect to all measures in column  $j$ ;
- $\bar{x}_i$  = a measurement in cell  $ij$  standardized with respect to all measures in row  $i$ ;
- $d_{ij}$  = the difference between the score in cell  $ij$  and the score in cell  $il$ ;
- $D$  = the distance between variables  $j$  and  $l$ ;
- $D(s)$  = this distance  $D$  found with  $\bar{x}$  scores in the matrix;
- $D(X)$  = this distance  $D$  found with raw scores in the matrix.

If the  $k$  dimensions are mutually orthogonal, the distance  $D$  between any two points fixed in this space is given by the square root of the sum of the squared differences between coordinates on the same dimension, i.e.,

$$D = \sqrt{\sum_i d_{ij}^2} \quad [1]$$

This distance,  $D$ , directly provides a measure of the relationship between variables  $j$  and  $l$ . For  $n$  variables one may then find  $n(n-1)/2$  values of  $D$ . Since  $D$  has a direct geometrical counterpart,<sup>3</sup> the  $n$  variables can be plotted in space such that the distances between each variable and every other variable satisfy the values in the matrix of  $D$ . Actual physical representation of such relations is limited to cases where  $k=3$ , but the values in the  $D$  matrix are accurate for  $n$  dimensions (however, see discussion below).

In Table 1 are given  $X$  and  $D(X)$  values derived from the hypothetical profiles drawn in Figure 1. The values for  $D$  indicate that variable A is less closely related to C than it is to B: ( $D(X)$  for A, C equals

TABLE 1  
RAW DATA ( $X$ ) AND DISTANCE MEASURES ( $D(X)$ ) FOR A SAMPLE OF VARIABLES  
(A, B, C, ETC.) LOCATED ON A SAMPLE OF SCALES

Scale	X					D(X)				
	A	B	C	D	E	A	B	C	D	E
a	1	2	4	6	7	A				
b	4	5	7	6	5	B	$\sqrt{5}$			
c	3	4	6	6	6	C	$\sqrt{45}$	$\sqrt{20}$		
d	2	3	5	6	7	D	$\sqrt{58}$	$\sqrt{31}$	$\sqrt{7}$	
e	4	5	7	6	5	E	$\sqrt{72}$	$\sqrt{45}$	$\sqrt{21}$	$\sqrt{4}$

$\sqrt{45}$  while  $D(X)$  for A, B equals  $\sqrt{5}$ ); the correlational measures are 1.00 in both cases. Profiles C, D, and E are drawn so as to have equal mean discrepancies with respect to A, yet so as to have radically different patterns of covariation with A. *Under these conditions, D takes into account profile relations.* This measure of relationship is minimal (e.g., the variables are closest together) when there is parallel covariation (for A, C,  $D(X)$  equals  $\sqrt{45}$ ); the distance increases for unrelated profiles (for A, D,  $D(X)$  equals  $\sqrt{58}$ ); the distance is maximal, and the indexed relationship least, when profiles have perfect negative correlation (for A, E,  $D(X)$  equals  $\sqrt{72}$ ). Thus  $D$ , as a measure of relationship, takes into account both the absolute discrepancy between sets of measurements as well as their profile similarities. The operations

<sup>3</sup> In our original work we had thought this method somehow dependent on correlational matrices. It was only after discovering that the method applied to raw score data that a mathematical rationale became apparent. We are grateful to Professor Jozef Cohen for pointing out the relationship between our  $D$  and the geometrical model.



for finding  $D$  have the additional advantage of being extremely simple and non-time-consuming. One simply sums the squared differences, one row at a time, for each pair of variables and looks up the square root of each such sum in the usual tables. Using calculating equipment, this is essentially a single operation upon the matrix.

### Relations to Factor Analysis

The existence of a geometrical model for the  $D$  matrix immediately suggests a possible relation to the factor analytic model. Suppose that the variables  $j$  and  $l$  are part of a "battery" of variables whose intercorrelations have been factor analysed, and suppose further that this analysis indicates two factors. The factor loadings are coordinates on the two factors which fix the two points in two-dimensional space. These two points represent the termini of two vectors whose lengths are  $h_j$  and  $h_l$  respectively. The relation between  $j$  and  $l$  can be expressed independently of the coordinates by the correlation  $r_{jl}$ , which in the factor analytic model is the scalar product  $h_j h_l \cos \phi_{jl}$ , where  $h$  is the square root of the communality and  $\phi_{jl}$  is the angle between the vectors  $j$  and  $l$ . By the law of cosines, the distance,  $V$  (to distinguish it from  $D$ ), between the termini of the two vectors  $j$  and  $l$  is found from the relation:

$$V = \sqrt{h_j^2 + h_l^2 - 2h_j h_l \cos \phi_{jl}} = \sqrt{h_j^2 + h_l^2 - 2r_{jl}} \quad [2]$$

If we now transform the raw score matrix ( $X$ ) to standardized scores  $z$  and compute the  $D_z$  matrix, we can write the following relations:

$$D(z) = \sqrt{\sum_i \left( \frac{z_j}{i} - \frac{z_l}{i} \right)^2} = \sqrt{\sum_i \frac{z_j^2}{i^2} + \sum_i \frac{z_l^2}{i^2} - 2 \sum_i \frac{z_j z_l}{i^2}} \quad [3]$$

Since,

$$\frac{\sum_i \frac{z_j^2}{i^2}}{k} = \frac{\sum_i \frac{z_l^2}{i^2}}{k} = 1.00, \quad \text{and} \quad \frac{\sum_i \frac{z_j z_l}{i^2}}{k} = r_{jl}, \quad [4], [5]$$

$$D(z) = \sqrt{k + k - 2kr_{jl}} = \sqrt{2k(1 - r_{jl})} \quad [6]$$

When

$$h_j^2 = h_l^2 = 1.00,$$

$$V = \sqrt{2(1 - r_{jl})} \quad 71$$

therefore,

$$D(z) = \sqrt{k} V. \quad [8]$$

Thus, under certain conditions—standardization with respect to all measures for the variables and all communalities equal to unity—a linear relation exists between distances in the  $D$  model and distances ( $V$ ) in the factor analytic model. It should be noted that the operation of standardizing measurements eliminates information regarding mean differences exactly as do the operations followed in computing product-moment correlations.

TABLE 2  
DATA OBTAINED FROM ONE SUBJECT JUDGING 10 CONCEPTS  
ON A SET OF 20 SEMANTIC SCALES

Semantic Scale	A QUICKSAND	B WHITE ROSE BUDS	C DEATH	D HERO	E METHODOLOGY	F FATE	G VIRILITY	H GENTLENESS	I SUCCESS	J SLEEP	
high	6	3	7	1	3	6	2	4	1	5	low
green	4	2	4	7	5	4	7	1	6	1	red
weak	7	1	6	7	5	5	7	1	7	3	strong
rough	5	7	2	3	2	2	3	7	5	7	smooth
active	7	7	7	1	3	7	1	7	1	7	passive
empty	7	5	1	7	5	1	7	7	7	6	full
small	7	1	5	7	2	4	7	4	7	5	large
cold	1	3	1	7	5	2	7	5	7	3	hot
clear	7	4	7	1	5	6	1	6	1	7	hazy
young	5	1	7	1	1	7	1	4	2	4	old
good	7	1	7	1	1	5	1	2	1	3	bad
peaceful	4	1	5	5	3	5	5	1	4	1	ferocious
sick	3	7	1	7	5	3	7	4	7	6	healthy
angular	7	7	3	6	3	5	6	7	5	7	rounded
tense	6	7	6	5	2	5	2	7	4	7	relaxed
sad	1	6	1	7	5	2	6	5	7	6	happy
soft	3	1	4	5	3	2	7	1	6	1	loud
wet	2	4	5	5	3	5	3	4	3	3	dry
beautiful	7	1	7	2	3	5	2	1	2	2	ugly
fresh	6	1	7	1	1	6	1	3	1	2	stale

*Application to Sample Data*

One subject rated 10 arbitrarily chosen concepts (such as HERO, FATE, GENTLENESS, and SLEEP) against the set of 20 seven-point scales shown in Table 2.  $D(X)$  values by the difference method as well as intercorrelations between all concepts were computed. The

TABLE 3

DISTANCE MEASURES ( $D(X)$ ) RELATING THE 10 CONCEPTS GIVEN IN TABLE 2

	A	B	C	D	E	F	G	H	I	J
A										
B	16.70									
C	9.32	19.10								
D	17.75	14.90	20.27							
E	15.30	11.45	16.13	10.25						
F	9.45	15.33	5.29	17.52	13.12					
G	17.70	16.73	20.40	4.36	9.90	17.83				
H	13.60	6.86	16.61	15.40	12.33	13.68	16.43			
I	17.40	14.56	20.22	3.74	9.95	17.64	3.87	14.80		
J	12.10	7.21	15.72	15.30	12.21	12.61	16.22	4.58	14.56	

TABLE 4

FACTOR LOADINGS AND COMMUNALITIES FOR THE VARIABLES GIVEN IN TABLE 2

Variable	Factor Loading			Communality
	I	II	III	$h^2$
A	-.44	.37	.42	.51
B	.59	.73	-.07	.93
C	-.93	.16	.29	.97
D	.81	-.43	.25	.90
E	.62	-.19	.16	.45
F	-.79	.31	.24	.78
G	.77	-.53	.26	.94
H	.37	.78	-.09	.83
I	.85	-.43	.21	.95
J	.35	.82	.16	.82

matrix of intercorrelations was factor analysed, yielding three factors. The raw score data are given in Table 2. The matrix for  $D(X)$ , obtained by the operation already described, is shown as Table 3. Table 4 gives the factor matrix and Table 5 the means and variances for the same

TABLE 5  
MEAN SCALE POSITIONS AND VARIANCES FOR THE VARIABLES  
GIVEN IN TABLE 2

Variable	Mean	$\sigma^2$
A	5.10	4.3
B	3.50	6.1
C	4.65	5.4
D	4.30	6.4
E	3.25	2.1
F	4.35	3.1
G	4.15	6.6
H	4.05	5.1
I	4.20	5.9
J	4.30	4.8

original data as in Table 2. Certain relationships are evident from inspection of Tables 3 ( $D(X)$  values) and 4 (factor loadings): variables having similar loadings on the three factors are close together in terms of the  $D$  distances (compare variables D, G and I).

Figure 3 is a drawing of the three-dimensional model constructed

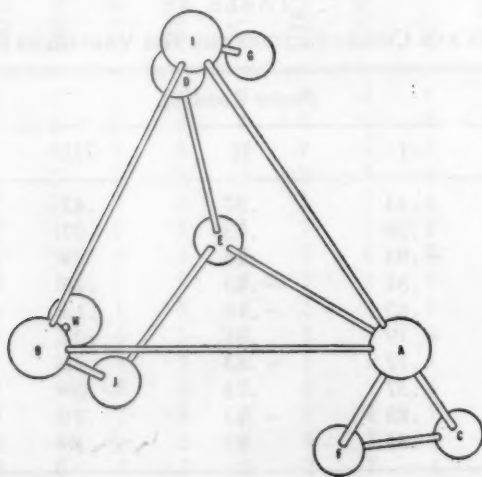


FIG. 3. DRAWING OF THREE-DIMENSIONAL MODEL CONSTRUCTED FROM DISTANCE MEASURES GIVEN IN TABLE 3.

VARIABLES D, G, I (HERO, VIRILITY, SUCCESS), B, H, J (WHITE ROSE BUDS, GENTLENESS, SLEEP), A, C, F (QUICKSAND, DEATH, FATE), AND E (METHODOLOGY).



from the  $D(X)$  values in Table 3. In constructing these models each  $D(X)$  value, as a distance on some convenient linear scale, is conceived as the radius of a sphere: variable B may be placed anywhere on a sphere 16.70 units from variable A; variable C must fall somewhere on the circle defined by the intersection of two spheres, 9.32 units from A and 19.10 units from B; variable D must fall at one of the two points of intersection of three spheres with radii of 17.75, 14.90, 20.27 from A, B, and C respectively. The position of variable F is completely determinate within three dimensions, as are the positions of all remaining variables. In practice we use small rubber balls to represent the variables and thin "pick-up" sticks to maintain correct distances. The model in Figure 3 specifies three clearly isolated clusters of variables: A, C, F (QUICKSAND, DEATH, and FATE), B, H, J (WHITE ROSE BUDS, GENTLENESS, and SLEEP), and D, G, I (HERO, VIRILITY, and SUCCESS). The remaining variable, E (METHODOLOGY) is independent of these clusters. The validity of this empirical categorization of these concepts is obvious.

Figure 4 shows the relation between  $D(X)$  distances and  $V$ , distances between the end points of vectors derived from the factor matrix for the same data, as a correlational scattergram. The close relation is apparent. The encircled points, falling somewhat away from the linear relation, all involve either variable A or variable E. Inspection of Table 4 indicates that these variables are most deviant from 1.00 in their values for  $h^2$ ; inspection of Table 5 indicates that these variables are also among those most deviant in terms of means and variances. This is an empirical demonstration of the relation between  $D(X)$  and  $V$  derived above, in this case employing unstandardized raw data. Although we have applied the difference method to relations among concepts in the present instance, the same procedure can be used to determine relations among the scales of judgment (tests) or among the subjects who do the judging.

#### SOME PROBLEMS CURRENTLY BEING INVESTIGATED

One problem appears in the illustration used above: Why, when 20 scales are used for judgments, do the concepts plot in three dimensions? It is certain that these 20 scales are not all independent;<sup>4</sup> they are oblique to varying degrees and may fall within a three-dimensional space. With oblique axes the distance formula proposed above probably distorts

<sup>4</sup> Factor analysis of 50 descriptive scales, from which these 20 were drawn, indicates that many of them are closely interrelated.

the "true" relation, but the nature and amount of such distortion is debatable. If  $x$  and  $y$  are two orthogonal axes and a set of points plotted on them falls in a straight line, then  $x$  and  $y$  are perfectly related, the straight line has reduced the dimensionality from two to one, and the relations among the points are not distorted. When conditions are not extreme—when dimensions are neither independent nor perfectly correlated—there may be distortion of an as yet undetermined magnitude. Distortion will definitely occur if one factor is represented by a large

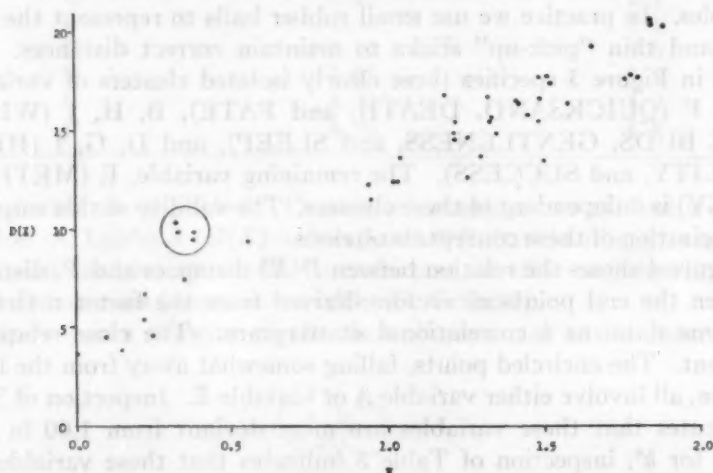


FIG. 4. SCATTERGRAM REPRESENTING CORRELATION BETWEEN  $D(X)$  VALUES (DISTANCES BETWEEN VARIABLES DETERMINED BY DIFFERENCE METHOD) AND  $V$  VALUES (DISTANCES BETWEEN END POINTS OF VECTORS REPRESENTING SAME VARIABLES IN FACTOR ANALYTIC METHOD).

THE ENCIRCLED POINTS ALL INVOLVE EITHER VARIABLE A OR VARIABLE E (SEE TEXT).

number of the  $k$  dimensions whereas the remaining dimensions are mutually independent. For the present, and for the purposes of our semantic measurements, we are solving these problems by finding a set of scales known to be independent from the results of factor analysis.

Another problem concerns the units of the  $k$  dimensions, which must be comparable if  $D$  distances are to be interpretable. In the case of our semantic measurements this is automatically taken care of by the format of the scales and instructions to the subjects. With other kinds of data, e.g., personality tests scores, the units may be different from measurement to measurement. The value  $D$  would be affected by this

inconstancy and the results liable to misinterpretation. In such cases it is possible to standardize or rank scores across variables for each measurement. The score matrix would then consist of  $z_i$  and the  $D$  matrix of  $Dz_i$ . The judgment of the investigator will determine when such transformations are necessary.

We have not as yet developed adequate methods for determining the significance of differences between distances. Although the clusters displayed in Figure 3 are self-evident, this is not always the case and we need a method for determining when a variable does or does not fall within a cluster. Similarly, in comparing the structures of variables obtained from an individual or group at different times (e.g., when an experimental factor such as psychotherapy has been introduced), we should like to be able to estimate the significance of changes in relationship. Chi-square and  $F$  seem the most promising statistics of those now available, but it may prove necessary to develop tests directly from the logic of this method itself.

Finally, there are two phenomena for which we have no adequate rationale, but which have empirical validity. First, simple summation of absolute differences,  $\sum_i |d_{ji}|/k$ , results in a matrix of distances which plots just as accurately as  $\sqrt{\sum_i d_{ji}^2}$ . Furthermore, these distances relate to factor analytic results just as well as do those in a  $D(X)$  matrix. Whether these results are artifacts of the kinds of numbers we have used or are due to some mathematical relation has not yet been determined. Second, when  $D$  is found from a matrix of intercorrelations, the results yield very accurate representations of the factor analytic pattern for the same variables. Using the full symmetrical matrix of correlations and entering the main diagonal with the highest correlations in each column (Thurstone's estimate of communality),  $D$ 's are obtained by exactly the same operation as described above. When  $V$ 's (from factor loadings) and  $D$ 's (from correlation matrix) were computed for Thurstone's box and cylinder problems,<sup>5</sup> the correlations between these two sets of distance estimates were .97 and .94 respectively.

#### SUMMARY

A method for analyzing interrelationships among variables has been described which takes into account not only the profile similarity

<sup>5</sup> THURSTONE, L. L. *Multiple-factor analysis*. Chicago: University of Chicago Press, 1947.

among the sets of measures but also their mean differences. This measure,  $D$ , can be applied to matrices of raw scores and, apparently, to correlational matrices as well. It has the advantage of an existing geometrical model which permits graphic representation of the data in many cases. It has the further advantage of unusual simplicity and speed in procedure. It has been shown to yield results identical to those of factor analysis under certain specified conditions. But the chief value of this method is not as an approximation to factor analysis, but as an independent method which does not eliminate information regarding differences between the means of the related variables. A number of unsolved problems are currently being studied.

*Received for priority publication March 5, 1952.*

### ERRATUM

An error occurred in the citing of various references in the article "Thematic Apperception Test: Interpretive Assumptions and Related Empirical Evidence," by Gardner Lindzey, in the January 1952 issue of *THIS JOURNAL*. The correction is as follows:

On page 1, the reference "Lawrence (39)" should read "Douglas (18)." With the exception of this reference, throughout the article author references numbered from 18 through 38 should have one number added to them to give the correct item in the reference list, e.g. Freud (24) should read Freud (25).



## A NOTE ON FUNCTIONAL RELATIONS OBTAINED FROM GROUP DATA

MURRAY SIDMAN

*Columbia University*

The empirical determination of functional relations between behavior and its controlling variables forms a large part of modern behavioral research. One important aspect of this type of experimentation is the method of distributing subjects among the various points which determine an empirical curve.

The most direct method is to use a single organism, and the same organism, to obtain every point on the curve. This procedure is not always practicable, however, for one or both of two reasons.

1. Intra-organism variability may be so great as to obscure any lawful relation. It is sometimes possible to avoid this problem by taking several determinations at each point and using a statistical measure, a common technique in obtaining threshold measurements (2).

2. Even this procedure will not be effective if, as is often the case, the experimental operations involved in determining one point on the curve have an effect upon the values of other points. For example, one cannot use the same organism to determine all the points on a function relating extinction responding to number of reinforcements. One reason for this is that the extinction operation is itself a variable entering into extinction results subsequent to reconditioning (4). It is seldom, if ever, possible to get around this difficulty by using a different individual for each point on the curve. Here inter-organism variability comes into the picture to obscure lawfulness.

Faced with these problems, most experimenters turn to group data. One technique is to employ the same group of organisms to obtain all the points. This procedure, however, is also ruled out if the second situation mentioned above is in effect (unless, of course, this is the problem under investigation). The only recourse remaining is to use a different group to determine each point. The rest of this paper will be devoted to a discussion of certain considerations involved in this latter method of obtaining an empirical function.

### INDIVIDUAL VS. AVERAGED FUNCTIONS

The first point to be made is that the mean curve obtained by such a procedure is not necessarily of the same shape as the inferred individual curves. (The term "inferred" is used here since this method is generally employed when the individual curves cannot be obtained directly.) The following development brings this out clearly. For the

purpose of demonstration we take as our example the negatively accelerated positive growth function which has achieved a certain prominence in behavior theory. This function can be expressed as

$$y = M - Me^{-kx}, \quad [1]$$

where  $M$  is the asymptote approached by  $y$ , and  $k$  determines the rate of approach to  $M$ . If the curves for individual organisms are of this shape, inter-organism variability might occur in the asymptotes ap-

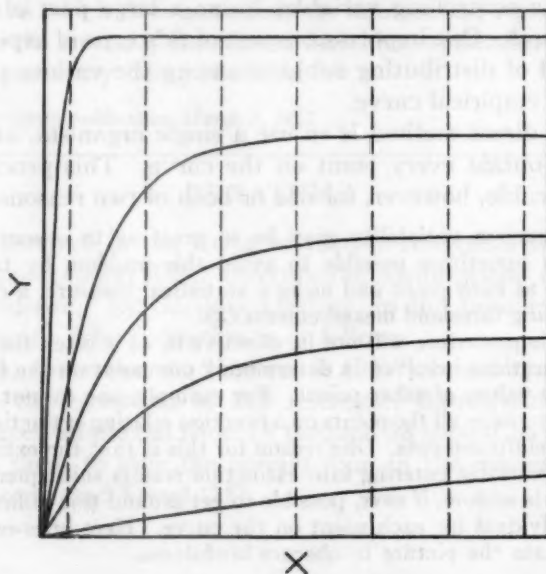


FIG. 1. SAMPLE SET OF INDIVIDUAL CURVES OF THE FORM,  $y = M - Me^{-kx}$ . EACH CURVE DIFFERS WITH RESPECT TO BOTH  $M$  AND  $k$ .

proached by the curves, in the rates of approach to the asymptotes, or both. Figure 1 represents a set of individual curves which vary with respect to both constants. (Although, for the sake of simplicity in Figure 1,  $M$  and  $k$  are assumed to be positively correlated, this assumption is not necessary.) When, for the reasons mentioned above, it is not possible to obtain these individual curves empirically, the procedure generally followed is to expose a different sample of the population of subjects to each value of the independent variable,  $x$ , and to take the mean of the dependent variable as the corresponding value of  $y$ . On the assumption that each of the samples is equally representative of the population, this procedure is represented in Fig. 1 by the broken

lines drawn from selected values of  $x$ . These lines simply indicate that, in a given experiment, the distribution of functions is cut through at selected values of the independent variable.

Corresponding to the curves of Figure 1, we can write the following equations:

$$\begin{aligned} y_1 &= M_1 - M_1 e^{-k_1 x} \\ y_2 &= M_2 - M_2 e^{-k_2 x} \\ &\vdots \\ y_n &= M_n - M_n e^{-k_n x}. \end{aligned} \quad [2]$$

To determine the mean value of  $y$  for a given value of  $x$  we sum equations [2] and divide by  $n$ , which results in the expression,

$$n^{-1} \sum_{i=1}^n y_i = \bar{y} = n^{-1} \left( \sum_{i=1}^n M_i - \sum_{i=1}^n M_i e^{-k_i x} \right), \quad [3]$$

which can be written

$$\bar{y} = n^{-1} \left[ \sum_{i=1}^n M_i - (M_1 e^{-k_1 x} + M_2 e^{-k_2 x} + \dots + M_n e^{-k_n x}) \right]. \quad [4]$$

Each of the exponentials in

$$S = M_1 e^{-k_1 x} + M_2 e^{-k_2 x} + \dots + M_n e^{-k_n x} \quad [5]$$

can be expanded to give

$$\begin{aligned} S &= [M_1 + M_1(-k_1)x + M_1(-k_1)^2 x^2/2! + M_1(-k_1)^3 x^3/3! + \dots] \\ &\quad + [M_2 + M_2(-k_2)x + M_2(-k_2)^2 x^2/2! + M_2(-k_2)^3 x^3/3! + \dots] \\ &\quad + \dots \\ &\quad + [M_n + M_n(-k_n)x + M_n(-k_n)^2 x^2/2! + M_n(-k_n)^3 x^3/3! + \dots]. \end{aligned} \quad [6]$$

Upon rearranging coefficients we have

$$\begin{aligned} S &= \sum_{i=1}^n M_i + [M_1(-k_1) + M_2(-k_2) + \dots + M_n(-k_n)]x \\ &\quad + [M_1(-k_1)^2 + M_2(-k_2)^2 + \dots + M_n(-k_n)^2]x^2/2! + \dots \\ &\quad + [M_1(-k_1)^m + M_2(-k_2)^m + \dots + M_n(-k_n)^m]x^m/m! + \dots. \end{aligned} \quad [7]$$

This can be expressed

$$S = \sum_{i=1}^n M_i + A_1 x + A_2 x^2/2! + \dots + A_m x^m/m! + \dots, \quad [8]$$

where

$$A_m = M_1(-k_1)^m + M_2(-k_2)^m + \cdots + M_n(-k_n)^m. \quad [9]$$

Substituting equation [8] into equation [4], we arrive at

$$\bar{y} = [1 - (1 + A_1x + A_2x^2/2! + \cdots + A_mx^m/m! + \cdots)]n^{-1}. \quad [10]$$

Equation [10] will reduce to the form [1] if and only if

$$A_i = A_j \quad [11]$$

for all  $i, j$ , or if

$$A_i = A_1^i. \quad [12]$$

Condition [11] is impossible, since the  $A$ 's are alternately negative and positive. Condition [12] is easily demonstrated to be impossible unless the  $A$ 's all equal unity, in which event [12] becomes a special case of [11].

It has been shown, then, that for individual curves of the form [1], if inter-organism variability occurs both in the asymptotes and in the rates of approach to these asymptotes, the average curve *cannot* be described by an equation of the form [1].<sup>1</sup> It can be seen from equations [7] to [10] that this will also be the case if the asymptotes are equal and variability occurs only in the rates of approach. Only when the rates of approach are equal will the mean curve be of the form [1]. Thus, under the assumption that

$$k_i = k_j \quad [13]$$

for all  $i, j$ , equation [4] can be rewritten

$$\bar{y} = n^{-1} \left( \sum_{i=1}^n M_i - \sum_{i=1}^n M_i e^{-kx} \right). \quad [14]$$

As far as this writer is aware, the assumption that the rates of approach to the asymptotes are equal for all the organisms in a given experiment has never been explicitly acknowledged by any experimenter or theorist who has fitted this growth function to data obtained by the method under discussion.<sup>2</sup>

At this point it may be argued that although the mean curve is

<sup>1</sup> The author is indebted to Mr. L. A. Gardner, Jr. for the essential elements of this demonstration.

<sup>2</sup> Hull appears actually to have made the opposite assumption. He states, "The 'constant' numerical values appearing in equations representing primary molar behavioral laws vary . . . from individual to individual . . ." (3, Postulate 18).



not the same as the individual curves, it is similar enough that, within the limits of experimental error, it can be fitted satisfactorily by the same function. Although this argument possesses dubious merit on grounds of theoretical consistency, it can also be attacked by demonstrating that many other types of individual curve will, if averaged, give as good an approximation to [1] as will equation [10].

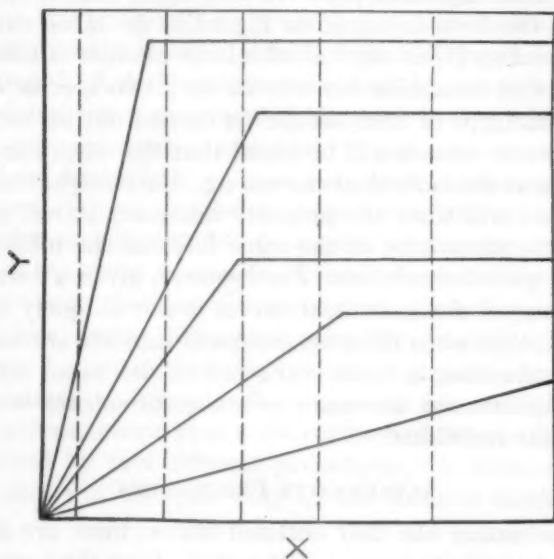


FIG. 2. SAMPLE SET OF INDIVIDUAL CURVES OF THE FORM  $y = mx$  UP TO A POINT OF DISCONTINUITY, AFTER WHICH  $y = c$ .

For example, if the individual curves are straight lines of the form

$$y = mx \quad [15]$$

up to a given value of  $x$ , at which point there occurs a discontinuity (see Fig. 2) after which

$$y = c, \quad [16]$$

it can be shown that the mean curve can be described by

$$\bar{y} = f(x)(1 - ax) + aBx, \quad [17]$$

where  $B$  is the sum of the maximum values of  $y$  and  $a$  is a proportionality constant between the slope,  $m$ , and the maximum value,  $y_c$ , of  $y$ . (This assumption of proportionality is not necessary, but is made merely

to simplify the discussion.)  $f(x)$  is a function describing the relationship between  $x$  and the sum of the  $y_e$  which have been reached at any value of  $x$ . (It can be seen from Fig. 2 that, as  $x$  increases, more of the individual curves will reach their maximum values of  $y_e$ .)  $f(x)$  will be determined by the frequency distribution of  $y_e$  and the relation, if any, between  $y_e$  and  $x$ . The form of  $f(x)$  will determine how closely equation [16] approximates equation [1]. We see, then, that if the individual curves are of the form indicated in Figure 2, the mean curve may approximate equation [1] or any one of a large number of other forms.

Although this discussion has treated only two specific cases in any detail, the same type of analysis can be carried out for any functional relation. In some cases it will be found that the mean curve will be of the same form as the individual curves, e.g., the straight line. However, many functions will show the property discussed above, namely, that the mean curve cannot be of the same form as the individual curves except under special conditions. Furthermore, given a particular mean curve, the form of the individual curves is not uniquely specified. It appears, then, that when different groups of subjects are used to obtain the points determining a functional relation, the mean curve does not provide the information necessary to make statements concerning the function for the individual.

#### ALTERNATIVE PROCEDURES

Given a situation like that outlined above, there are several alternatives open to experimenters and theorists. First the suggestion might be made that all data obtained by the averaging procedure outlined above be ignored and that the questions which such data attempt to answer not be asked. This radical solution is probably not necessary. Such mean curves may give some valuable information, depending upon the validity of the assumptions one is willing to make concerning the general lawfulness of individual behavior. If it is assumed that all individuals of a certain class will display the same type of functional relation in a given situation, then the mean curve will tell something about that function. If, under this assumption, we obtain the mean curve described by equation [10], it will be known that the individual curves are increasing functions of  $x$  and that they either reach a maximum or approach an asymptote.

However, we are not forced to make such an assumption. The mean curve of equation [10], for instance, can be obtained even if the individual curves are so irregular that they cannot be described by any useful equation. A more profitable approach might be to obtain

and present all data in the form of distributions, to specify the distributions by their form and by their parameters, and to relate these distributions to the independent variables. (Such procedures would, of course, apply also to experiments in which the same group of subjects is used to determine all points on a function, but where it is observed that the individual data are not amenable to a functional description.)

A third alternative is to develop techniques which will produce lawful individual functions and to present the data without averaging. Although many methods have already been developed for work with individuals (e.g. 1, 5, 6, 7), many more would have to be worked out either by devising new measuring techniques or by attaining more rigorous control over extraneous variables. Statistical procedures would enter into these methods in at least two ways. Replicative statistics would be necessary to determine the reliability of the individual curves, and information would probably be needed concerning the population distribution of the curve constants. Even if these methods were highly developed, however, there are still some data (such as speed of acquisition of behavior under different motivating conditions) which will probably never be amenable to individual treatment. It will, in such cases, be necessary to forego such data or to use statistical analysis. If the latter is done there remains the problem of theoretical integration of data obtained by two different procedures. A decision among the alternate choices will be made only on the basis of further empirical investigation.

## REFERENCES

1. FRICK, F. C., & MILLER, G. A. A statistical description of operant conditioning. *Amer. J. Psychol.*, 1951, **64**, 20-36.
2. GRAHAM, C. H., & VENIAR, F. A. The influence of size of test field surround on visual intensity discrimination. *Proc. nat. Acad. Sci.*, 1950, **36**, 17-25.
3. HULL, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173-180.
4. PERKINS, C. C., JR., & CACIOPPO, A. J. The effect of intermittent reinforcement on the change in extinction rate following successive reconditionings. *J. exp. Psychol.*, 1950, **40**, 794-801.
5. SKINNER, B. F. *The behavior of organisms: an experimental analysis*. New York: Appleton-Century, 1938.
6. SKINNER, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, **57**, 193-216.
7. SKINNER, B. F., & CAMPBELL, S. L. An automatic shocking-grid apparatus for continuous use. *J. comp. physiol. Psychol.*, 1947, **40**, 305-307.

Received August 20, 1951.

## A NOTE ON THE COMBINATION OF RATINGS ON THE BASIS OF RELIABILITY

C. H. LAWSHE AND BRYANT F. NAGLE

*Occupational Research Center, Purdue University*

In psychological research one is often faced with the question of how to combine ratings when a measure of rater reliability is the only attribute of the ratings known. This situation arises frequently in setting up a criterion when there is no way in which to determine the validity of the various ratings. The question becomes one of how to combine the ratings of the different judges on the basis of their reliability.

In partial answer to this problem, Shen (9) developed a formula for estimating the reliability of an individual rater from the intercorrelation matrix of the raters. Kelley (2) has followed up Shen's work by producing a table of weights that may be used when combining the ratings. With these tools, psychologists have been able to go in at least four directions: (a) weight the raters equally; (b) eliminate a rater with low reliability and weight the remaining ones equally; (c) weight the raters according to the weights given by Kelley; and (d) eliminate a rater with low reliability and weight the remaining ones according to their Kelley weights. This paper represents an empirical look at the results these four approaches will give.

The data used in comparing the methods come from actual studies. The five cases used are briefly described:

**Case 1.** Six industrial executives rated 46 general foremen in a tractor plant on the basis of how well the men were doing their jobs. The type of rating used was paired comparison, utilizing partial pairings as suggested by McCormick and Bachus (4) and McCormick and Roberts (5). From Nagle (8).

**Case 2.** Five supervisors rated 14 office jobs on the basis of which should carry the highest pay rate. Paired comparison ratings were used. From Miles (7).

**Case 3.** Four professors rated 23 graduate students in industrial psychology on their ability to do graduate work. Paired comparison was used. From Lawshe and Deutsch (3).

**Case 4.** Four supervisors rated 24 journeyman proofreaders on their versatility in proofreading. Paired comparison was used. From McGinley (6).

**Case 5.** Three graduate students in psychology rated 40 responses to a picture on a positive-negative attitude expressed towards the theme represented by the picture. A five-point rating scale was used. From Boyce (1).

Data drawn from these five cases are presented in Table 1. In columns 2 and 3 the number of raters and the number of rated objects are shown. Shen's formula was applied to obtain a reliability estimate for each rater, and the estimate for the least reliable judge is presented

in column 4. Column 5 shows the mean Shen reliability<sup>1</sup> of all judges in the case.

As indicated above, one approach involves the computation of the mean of all the possible intercorrelations between the judges and the subsequent stepping up of this mean by the Spearman-Brown formula.

TABLE 1

ESTIMATED RELIABILITY OF RATINGS WITH VARIOUS NUMBERS OF JUDGES USING EQUAL AND UNEQUAL WEIGHTINGS

Case	No. Raters	No. Things Rated	Lowest Shen Reliability	Average Shen Reliability	Ave. Intercorrel. of Judges		Spearman-Brown Reliability		Weighted Composite Reliability	
					N	N-1	N	N-1	N	N-1
					Judges	Judges	Judges	Judges	Judges	Judges
1	2	3	4	5	6	7	8	9	10	11
1	6	46	.304	.57	.549	.605	.880	.884	.890	.887
2	5	14	.568	.79	.800	.860	.952	.961	.955	.972
3	4	23	.394	.58	.585	.643	.849	.844	.860	.852
4	4	24	.695	.91	.825	.862	.950	.949	.986	.951
5	3	40	.189	.44	.400	.543	.667	.704	.701	*

\* Shen reliabilities cannot be computed where there are only two raters.

The resulting statistic yields an estimate of the pooled ratings of all judges with individual ratings equally weighted. Column 6 shows the average of all of the intercorrelations for each case and column 8 shows the corresponding stepped up values.

If, on the other hand, we do not use all of the judges but in each case drop out the single judge with the lowest Shen estimate (the one whose reliability is reported in column 4) and compute the mean intercorrelation of the remainder, we obtain slightly higher values. These values are reported in column 7 in the table. When these ratings are pooled (weighted equally) the reliability of the composite may be estimated by the Spearman-Brown values as shown in column 9.

Kelley's recommended procedure (weighting ratings *unequally* in relation to reliability) yields composite ratings, the reliability of which may be estimated by means of the following formula:<sup>2</sup>

<sup>1</sup> Computed after conversion to Fisher's *s*.

<sup>2</sup> For this formula the authors are indebted to Mr. James Norton, Division of Educational Reference, Purdue University. It was derived from the formulas used by Kelley to determine weights for reliability coefficients.



$$r_w = \frac{\sum r_i W_i^2 + 2 \sum r_{ij} W_i W_j}{\sum W_i^2 + 2 \sum r_{ij} W_i W_j}$$

where  $r_i$  = Shen reliability of any given rater  $i$ ,

$W_i$  = Kelley weight for rater  $i$ ,

$W_j$  = Kelley weight of any rater except  $i$ ,

$r_{ij}$  = correlation between rater  $i$  and rater  $j$ .

If all raters have equal Shen reliabilities, this formula will give the same composite reliability estimate as the Spearman-Brown formula.<sup>3</sup> If raters have differing reliabilities and consequently different weights this formula will always yield a higher estimate than the Spearman-Brown. It should be emphasized that the two estimates cannot legitimately be compared statistically because they are based on different assumptions. Column 10 contains the Shen-Kelley estimates using all raters, whereas entries in column 11 were obtained by eliminating the rater with the lowest reliability (i.e., the one listed in column 4), recomputing rater reliabilities and weights, and using the above formula.

The comparisons that should be made from this table are column 8 with 9, 10 with 11, 8 with 10, and 9 with 11. The first two comparisons (8 with 9, and 10 with 11) show the effect of eliminating a rater whose reliability is considerably below the average of the others. It will be noted that in the first four cases the differences between the reliabilities of  $N$  and  $N-1$  (columns 8 and 9) are slight, and there is no consistent tendency for the  $N-1$  reliabilities to be higher or lower than those of  $N$ . Case 5 shows a larger difference. No technique is known for estimating the statistical significance of this difference, but from a practical significance point of view it must be admitted that the difference is small.

A comparison between columns 8 and 10, and between 9 and 11 suffers from the fact that the computation of the entries in columns 8 and 9 employ different assumptions from those in 10 and 11. But some sort of comparison, "with reservations," can be made based on the fact that the Spearman-Brown procedure and the weighted composite reliability formula do yield the same results when raters are weighted equally and that when not equally weighted, the latter procedure results in a larger estimate. Inspection of the coefficients shows that the differences

<sup>3</sup> Since Fisher's  $z$  transformation should not be used in computing Shen reliabilities this is true only when Fisher's  $z$  transformation is *not* used in computing the average of the matrix of rater intercorrelations. The authors have employed Fisher's  $z$  in obtaining the values in columns 6, 7, 8, and 9. The differences between the values in columns 8 and 9 and those which would have been obtained had Fisher's  $z$  transformation not been used range from .000 to .013, the median being .003.

between the reliabilities are very small, none exceeding .036, though all the weighted composite reliabilities (columns 10 and 11) are larger as expected.

Within the limitations of the comparison made here, it can be seen that the decision to eliminate a low-reliability rater from a rater pool or to apply differential weights to different ratings is not too fruitful. The authors are forced to conclude that much of the time and effort spent in determining the reliabilities of raters and combining their ratings using differential weights is for naught. Even where there are wide differences in the reliabilities of the raters, the elimination of the very low rater does not seem to improve the composite reliability. In fact, the composite reliability may on occasion be lowered! When weights based on the reliabilities are applied to the ratings, no practical improvement in composite reliability appears to be effected.

Because we do not wish to be accused of taking a "pot shot" at some of our more statistically sophisticated colleagues, one further comment seems desirable. The statistician's job is one of finding new ways of making evaluations, of seeking more sensitive tests of significance, and of finding more nearly optimum ways of weighting and combining data. The applied psychologist must, however, reserve the prerogative of deciding how much refinement is enough. There is probably no substitute for empirical evidence in deciding whether or not a particular procedure is more precise than the nature of the data can justify.

## REFERENCES

1. BOYCE, J. E. Reliability of ratings of picture responses. Unpublished master's thesis, Purdue Univer., 1951.
2. KELLEY, T. L. *Fundamentals of statistics*. Cambridge: Harvard Univer. Press, 1947, 419-425.
3. LAWSHE, C. H., & DEUTSCH, S. The interests of industrial psychology students. *J. appl. Psychol.*, 1952, **36**, in press.
4. MCCORMICK, E. J., & BACCHUS, J. A. Paired comparison ratings: 1. The effect on ratings of reduction in the number of pairs. *J. appl. Psychol.* (accepted).
5. MCCORMICK, E. J., & ROBERTS, W. K. Paired comparison ratings: 2. The reliability of ratings based on partial pairings. *J. appl. Psychol.*, in press.
6. MCGINLEY, A. D. The investigation of criteria of job performance in a printing plant. Unpublished master's thesis, Purdue Univer., 1950.
7. MILES, MINNIE C. Studies in job evaluation: 9. Validity of a check list for office jobs. *J. appl. Psychol.*, 1952, **36**, 97-103.
8. NAGLE, B. F. A study of foreman responsibility. Unpublished master's thesis, Purdue Univer., 1951.
9. SHEN, E. The reliability coefficient of personal ratings. *J. educ. Psychol.*, 1925, **16**, 232-236.

Received August 18, 1951.

## BOOK REVIEWS

TOLMAN, EDWARD CHACE. *Collected papers in psychology*. University of California Press, 1951. Pp. xiv+269. \$4.50.

The bibliography at the end of this volume lists 82 titles, dating from 1917 to 1949. From this list an anonymous group of colleagues and former students have selected nineteen representative items for republication. The volume begins with "A New Formula for Behaviorism" (1922) and ends with "Cognitive Maps in Rats and Men" (1948). It is prefaced by a brief but excellently written appreciation of Tolman as scientist and as teacher. Tolman himself had nothing to do either with its initiation or with the selection. His only direct contribution to the book, apart from its content, is a footnote to the Foreword, in which he characteristically expresses the wish "that a mere half of the kind things said had been true." On this point the editors need no reassurance. They might have said many more kind words without violating the traditions of academic restraint. They deserve our gratitude, not merely for having paid fitting tribute to one of the great men of contemporary psychology, but also for having presented a volume which in itself is an important contribution to the history of recent psychological thought.

For more than thirty years Tolman has been in the thick of the battle, a battle in which it is sometimes difficult to tell who is fighting against whom, but in which the participants, especially Tolman, thrust and slash at each other with every evidence of enjoyment. The editors were probably wise in weighting the selection with polemical items, for, although Tolman has contributed generously to our store of psychological facts, he is at his liveliest when he marshals his forces—his rats, his neologisms and his amazing diagrams—and charges into the debate. For a convinced pacifist he has been involved in an extraordinary number of fights. It is characteristic, however, that he always apologizes before he knocks his opponent down and apologizes again as he helps him to his feet.

This book might well serve as "core" reading for a course in Systematic Psychology, for it gives us the history of a system coming into being and grappling in turn with virtually all the fundamental problems with which a psychological system must deal. Whether or not one agrees with Tolman, it is a refreshing and enlightening experience to look at the problems of psychology through Tolman's eyes. The reviewer will perhaps be pardoned if, instead of reviving the old controversies, he selects at random a few of the impressions with which the book has left him.

1. Tolman is not ashamed of, or afraid of, our philosophical heritage. For

him the persistent problems of psychology are the problems of cognition and of motivation. In the last analysis these are the problems of epistemology and of ethics. Tolman has battled stoutly for what he considers to be an objective and scientific psychology, but he has always considered his particular science as merely one of many approaches to an understanding of the totality of things. Tolman speaks to, and listens to, the philosophers. The rest of us might be better psychologists if we followed his example.

2. Much of Tolman's systematic writing has been in the form of protest; but behind every protest is an affirmation. He protested against the "mentalists," against the Titchenerian introspectionists, against teleology. He took issue with Watson and the muscle-twitchers, with Kuo and the anti-hereditarians, and with those who saw in the conditioned reflex a magic formula that would solve all psychological problems. His constant affirmation has been that psychology can and must be an "objective" science. The criterion of objectivity favors behavior rather than experience as the primary subject-matter of psychology, but it does not lead him to deny or even to disparage immediate experience. In fact, he asserts, as a phenomenologist would, that immediate experience "contains quite as much objectivity as it does subjectivity" and, as Titchener would, that it is "an initial common matrix out of which both physics and psychology are evolved" (p. 96). His behaviorism reflects the belief that, if psychology is to be a science, its facts and laws must be stated in a language that is not uniquely related to the experience of any one individual. Where Tolman differs from many other behaviorists is in his stubborn insistence that this language be rich enough to encompass every significant variable in behavior (and experience). He wants his behaviorism to deal with "real life," not just with *S-R* artifacts. Purpose, for example, is a fact of observation; therefore we must have words like means-end-readiness and sign-gestalt-expectation. This reviewer thinks he has a fairly good reading knowledge of the language, but he confesses that he is sure of its meaning only when he has found a counterpart for each word in his own experience.

3. Tolman is not one to retreat before a new idea, or an old one. He advances to meet it, with a gleam in his eye and another diagram to put on the board. In his early writings he was worried about the problems set by the psychologists of an older generation, by James, Titchener, McDougall, and Watson. Then he went to Vienna, and one has the feeling that there he really enjoyed himself. On the one hand, there were the Vienna positivists, who were thinking operationally, and there was Egon Brunswik, whom Tolman later welcomed into his laboratory in Berkeley. On the other hand, there was the lingering tradition of Brentano and act psychology. In all his subsequent writings one can recognize both types of operational emphasis—on the operational criterion in scientific definition, on the operation (act) as the subject-matter of psychology. Tolman's system has grown to meet the challenge of the Gestaltists and of Kurt Lewin. As the Freudian psychology gained in popularity he did not by-pass it like so many of his colleagues. He accepted the challenge and tried to find a place in operational behaviorism for the dynamics of the unconscious. (The diagrams at this point become quite bewildering.) If the book had continued beyond 1948, the reader would have had a chance to observe Tolman grappling with the fundamental concepts of sociology.



4. Probably no psychologist has ever learned so much from rats. Tolman loves them, he is really curious about them, and he believes that the essential variables of human behavior are to be found in the rat. He intends, as he says on p. 163, "to go ahead imagining how, *if I were a rat*, I would behave . . . ." But does the rat really furnish an adequate paradigm for the solution of the human riddle? Being a modest man, Tolman might admit that human behavior presents a greater variety of psychological phenomena; but he would probably argue that a sound training in rat psychology enables one to see the human problems more clearly. Perhaps he is right. To this reader it seems clear, however, that Tolman could never have been such a good rat psychologist had he not been a good human psychologist to begin with. The myopic, anomic behaviorist of the Watsonian era looked at the rat and saw nothing but muscle twitches. Tolman looked at the rat and saw much more, and animal psychology has been the richer for it.

One minor criticism, and a very friendly one. A volume of "Collected Papers" faintly suggests that a career has come to an end. How about a subtitle: "Tolman in Midpassage"?

ROBERT B. MACLEOD.

*Cornell University.*

KATONA, GEORGE. *Psychological analysis of economic behavior*. New York: McGraw-Hill, 1951. Pp. ix+347. \$5.00.

The general thesis of this book is that economic behavior can only be understood, predicted, and controlled by utilizing psychological methods of research and analysis. Like all social behavior, economic behavior is influenced by a variety of attitudes, habits, and motives (not solely by profit seeking); it takes into account the future ("expectations") as well as the past; it is characterized by impulsive and routine behavior in addition to genuine decision-making; and the knowledge, beliefs, attitudes, and habits that influence it are acquired and change both as a result of repetition and as a result of insights based on reorganizations of field.

The author devotes most of this book to raising psychological questions about the saving and spending decisions of consumers, the output, price, and investment decisions of businessmen and the attitudes and decisions of both consumers and businessmen under the special circumstances of inflation and wide fluctuations in business activity. Tentative answers to some of these questions are suggested on the basis of empirical evidence from surveys carried out by the Survey Research Center (and its predecessor agency).

This is an interesting and valuable book. It will be very useful for teaching purposes, and both psychologists and economists will benefit from its incisive demonstration of the gaps that exist in current economic materials—gaps which only psychological investigations are likely to fill. In the place of the good but vague intentions to "integrate

psyco  
for a  
of s  
pron  
and  
year  
T  
caut  
eval  
anal  
ment  
sugg  
beha  
sume  
T  
econ  
cepts  
of co  
other  
the b  
demo  
of the  
the s  
the o  
"This  
study  
The a  
are se  
from  
M  
psych  
The b  
behav  
tions  
indivi  
condi  
and g  
which  
influe  
term  
(other  
they n  
extens  
psyche  
does n



psychology and economics" found spasmodically in economic literature for a generation now, this book charts in realistic detail a wide variety of specific problems, investigation of which by psychological methods promises to contribute greatly to economic understanding, prediction and control. No doubt many doctoral dissertations over the next ten years will find their starting points in this book.

To the reader whose hopes will be aroused by the title, a word of caution is needed, however. The book does not attempt a general evaluation and synthesis of research significant for the "psychological analysis of economic behavior." Rather, it is intentionally a development of an individual viewpoint, and comprises primarily a series of *suggestions for research* on the psychology of *certain kinds* of economic behavior, illustrated by analyses of *survey data* (mainly from consumers), collected by *one research organization*.

The author devotes most of his book to a stimulating inquisition of economic dogma—raising questions on the basis of psychological concepts, analogies, case studies, and survey results. Although a wealth of consumer survey data is drawn upon, the comparative paucity of other reliable psychoeconomic data today unfortunately tends to make the book seem negative, for the author generally fails to replace the demonstrably unsatisfactory with the more satisfactory. Characteristic of the book is the refrain of nearly every section that "On the basis of the studies conducted up to now, we cannot foretell in advance when the one and when the other situation is likely to occur" (p. 95), and "This may be the case under certain conditions that require further study, but need not be the case under other conditions" (p. 167). The author speaks of "the present stage of our discussions, when we are searching for a program of research" (p. 137). This results, of course, from the author's deliberate choice of the inductive method.

Many readers will question the author's selection of the kinds of psychological problems and concepts deserving greatest attention. The book abounds in references to the prevalence of habit in economic behavior except where a reorganization of field provides the preconditions for a "genuine decision." Little attention is given, however, to individual variations in *knowledge* about economic trends and market conditions, as a major factor conditioning both habitual responses and genuine decisions, or to the channels of communication through which knowledge is acquired and attitudes and expectations are influenced. There is some discussion of levels of aspiration, and the term "attitude" is often used, but how aspirations and attitudes (other than expectations) arise, how they change and how closely they may be related to action is almost entirely neglected, despite the extensive empirical work carried on by others in this field. Likewise, psychoanalytical concepts and research get short shrift. The author does not consider that existing psychoanalytic concepts are very help-

ful in analyzing economic behavior, or that they are likely ever to be very helpful in understanding the "regularities of behavior that are responsible for relationships between, say, income changes and saving or for cyclical fluctuations in the present American economy" (pp. 58-59). He apparently does not consider (as the reviewer does) that, for example, anxiety feelings may significantly influence consumer or business behavior (e.g., during war or in business crises) or that ego-satisfaction, anxiety-strengthened prejudices, or a search for security may have important economic results by influencing top officials in strategic companies (e.g., price leaders) in strategic industries (those whose decisions have extensive repercussions on economic development or on business fluctuations).

The author also omits certain major areas of economics. He explicitly excludes employee behavior (aptitudes, incentives, training, technological change, mobility, organizations, etc.) and the whole field of international economics (including the burgeoning new field of international investment, technological diffusion, and governmental aid programs). Moreover, although he points out that policy makers in government need surveys to help them make decisions, he does not discuss the psychological factors that may affect governmental economic activities (public services, government-owned enterprises, regulatory agencies, fiscal management, legislation on economic matters, etc.) This omission is particularly noticeable in the discussion of inflation.

Conspicuously absent in the discussion of consumer behavior is any consideration of the psychological determinants of value, of the shape of indifference curves, and of the elasticity of demand. Despite these omissions, however, the two chapters on consumer spending and saving are the strongest in the book, with significant theoretical questions raised and pertinent empirical data adduced in hard-hitting fashion.

Psychological aspects of social interaction, communication and methods of social control (advertising, education, publicity, propaganda), in their bearing upon economic behavior, are also relatively neglected in the book.

The discussion of business behavior leans too heavily, in the reviewer's opinion, on data from small enterprises. It consequently probably overstates the extent of habitual behavior in business, for small companies have fewer facilities, and perhaps less pressure, to collect pertinent information, utilize a variety of specialists, perceive and review alternatives, and otherwise engage in the processes characteristic of genuine decision-making. Moreover, it is particularly among large firms that the phenomena of price-leadership and industry-dominance occur, and that highly personal decisions may have quite significant impacts on the economy. The book also gives much less emphasis than is deserved to the relationship between business behavior and the characteristics of the industry (average size of concerns, extent of public interest, stage of growth, competitiveness, nature of product, importance of labor costs, market served, etc.) and of the concern (structure, locus of control, relative size, etc.). The discussions of business investment (drawing heavily on Terborgh's work) and of retail behavior during inflation are, however, particularly effective.

In the reviewer's opinion, the book would have gained a good deal from inclusion of empirical materials in addition to data from surveys with which the author himself has been associated. No use is made of the extensive research

can  
Cen  
gran  
sti  
ratio  
ship  
cuss  
cons  
ness  
Com  
are  
5  
to g  
of th  
inve  
buyi  
for a  
indu  
Non  
Rese

7  
one

1  
for m  
bring  
rou  
effect  
prom  
actua

2  
differ  
expect  
sume  
Surv  
amou  
of ch

3  
with  
place  
tion a

A  
think  
signi  
despi  
psych

U

carried on by non-governmental organizations (other than the Survey Research Center) concerning consumer behavior and its relationship to income, geographic area, sex, race, size of town, education, reading habits, advertising stimuli, product changes, etc. There is no reference to the Psychological Corporation, for example, nor to the work of such men as Alderson, Britt, Blankenship, Link, Stanton, or Lazarsfeld (except in methodological notes). In discussing consumer and business attitudes and behavior during business cycles, consumer surveys are quoted, but no data are utilized from the surveys of business expectations made by *Fortune*, McGraw-Hill, the Securities Exchange Commission-Department of Commerce, and Dun and Bradstreet (though these are mentioned in the later chapter on methodology).

Similarly, many kinds of aggregate economic data might have been utilized to give point to the author's discussions of economic behavior, especially in view of the relative paucity of actual surveys of individual entrepreneurs. Data on inventory fluctuations, internal and external corporate financing, installment buying and other consumer credit, durable goods expenditures, expenditures for advertising outlays on plant and equipment, output and prices in different industries, etc., might well have been used when these subjects were discussed. None of the statistical series developed by the National Bureau of Economic Research or by the Department of Commerce is presented, however.

This book, therefore, is not nearly so general as its title would lead one to believe. In positive terms, what contribution does it make?

1. It presents what is by now a familiar, though generally unheeded, plea for more empirical studies of economic behavior, utilizing methods that will bring out the individual differences in motivations, expectations and extent of routine or habitual behavior involved. This plea is made more than usually effective by illustrations of assumptions or generalized statements made by prominent economists which are cast in doubt or even directly contradicted by actual survey data.

2. A good deal of material is presented to indicate the relative importance of different factors (past income, expected income, liquid asset holdings, price expectations, etc.), in influencing the spending and saving behavior of consumers. Although this material is in general limited to that collected by the Survey Research Center (and its predecessor agency), it is substantial in amount, of fairly wide scope, and unusually excellent in quality—both in terms of choice of pertinent subjects and in terms of survey technique.

3. A wide variety of questions are raised about the behavior of businessmen, with enough in the way of evidence and analogy to unsettle much of the complacency of economists (and of policy makers) about the simplicity of motivation and of other determinants of business behavior.

As a book that breaks through traditional ways of professional thinking and poses a multitude of new research problems of great significance, both for economists and for social psychology, this will, despite its incompleteness, be a landmark in the frontier discipline of psychoeconomics.

SAMUEL P. HAYES, JR.

*United States Special Technical and Economic Mission to Indonesia.*

ROHRER, JOHN H., & SHERIF, M. *Social psychology at the crossroads*. Pp. viii+437. New York: Harper, 1951. \$4.00.

This book reports the five-day conference on "Social Psychology at the Crossroads" held in 1950 at the University of Oklahoma. Social psychology is made visible in a new geographic center—both by the holding of the meetings and the publication of this volume. The book itself consists of summaries by well-established researchers. Each chapter relates a segment of concepts and theories which the contributor feels has relevance to the development of social psychology. But the volume contains no news of significant, hitherto unreported, advances.

To the reader looking for a summary of recent developments in fields related to social psychology, this volume will be useful. The materials presented are now available only in scattered sources. One familiar with the literature, however, could predict what most of the speakers talked about and what they would have to say. MacLeod, of course, wrote about phenomenology. Herskovits treated cultural vs. psychological reality. Barker and Wright illustrated the methods they use in "psychological ecology." Arensberg wrote about interactionism. And Harry Harlow, in talking about social learning in primates, ended with a plug for "learning set" theory. The contributions, even though often repetitious of each author's earlier writings, present a valuable overview of concepts and theory.

One expects variation in quality of contribution among sixteen contributors in a conference. This book meets that expectation. Some of the reports are carefully drawn documents, summarizing in an explicit and dramatic way the recent thinking and work of their authors. Postman's forging of a general theory of cognition and Schneirla's expansion of his "levels" concept of social systems are examples of this kind of document. Other contributors were apparently caught within the press of duties and wrote merely "off-the-cuff" reports.

Only four of the sixteen contributions were on social psychology per se. The other twelve were concerned with related disciplines. The volume is therefore appropriately entitled, "*Social Psychology at the Crossroads*" (italics mine). Rohrer and Sherif exhibit the lines of cross-traffic which are influencing social psychology, covering disciplines from genetics through anthropology. There is breadth in the coverage. For instance, one of the comparative psychologists (Schneirla) treats "individual properties which promote group formation." A few chapters later, an applied sociologist (Whyte) stresses the relevance of hierarchical organizations to the formulation of a social psychology.

There was duplication—and even triplication—within the conference which seems unwarranted, especially when one notes important omissions in making the junction at the crossroads cosmopolitan. For



instance, the psychiatrists were absent, even though many of them become increasingly interested in the role of the patient's social milieu. Perhaps Stanton and Schwartz, who are making observational studies of communication and interaction in hospital ward situations, would have been worth including.

There was not much effort devoted by the contributors to the particular ways in which their various disciplines *should* influence the development of social psychology. The comparative and experimental psychologists were more aware of the difference between development of their basic concepts and their subsequent application in social psychology. For instance, Volkmann was quite conscious of dangers in his reckless extrapolation of his work on anchoring points in perceptual judgment as applicable to the social psychology of education. The sociologists and anthropologists, on the other hand, took "social psychology" in their stride. They were quite unconcerned about differentiating their main focus of interest from the area being claimed by the embryonic social psychologists.

The four social psychologists represented in the volume are in essential agreement with each other in staking their claim. They posit that social psychology is the systemization of "how psychological processes function under the field conditions of group life," as Newcomb put it. And living at the crossroads provides a sound habitat if Hartley, Sherif, Sargeant, and Newcomb's reports are compared with the social psychology of the 1920's, when Allport and others conceived it more narrowly in the main stream of individual psychology.

Newcomb and Sargeant emphasize in their rich, theoretical chapters the inner life of the social individual. They stress "self-concepts" and social attitudes as generated and conditioned by group roles and memberships. This concentration meant omission in the conference of another emphasis in social psychology—understanding the behavioral life of social beings in groups as stressed in the work, for example, of Bales, Carter, or Festinger, who are concerned with the description of group functioning as related to the psychological nature of man. Sherif's chapter on "A Preliminary Experimental Study of Inter-Group Relations" takes this counterbalancing position. This last chapter of the book, however, is largely a well-written recital of findings in a particular experiment and fails to counteract the theoretical emphasis of Hartley, Sargeant, and Newcomb for the volume as a whole. This omission injures the catholicity of the book.

In general, the volume ably achieves its goal of reporting the University of Oklahoma's conference on "Social Psychology at the Crossroads." The book largely consists of recapitulations of theory and facts in many fields which have bearing upon the development of social psychology. It also includes two excellent summaries of current



theories in social psychology, conceived as the impact of social phenomena upon the psychological functioning of human beings.

HAROLD GUETZKOW.

*Graduate School of Industrial Administration, Carnegie Institute of Technology.*

QUEENER, E. L. *Introduction to social psychology*. New York: Sloane, 1951. Pp. xiv+493. \$5.00.

This text is somewhat different from others in social psychology in both content and organization. The trend toward greater use of anthropological and sociological material is pushed further than is yet customary. One objective—the understanding and prediction of interpersonal behavior—determines what data are selected from these two areas and from social psychology, and how they are organized.

The author makes the plausible assumption that it would be helpful, for predicting the behavior of any one person, to know: the characteristics of (1) human beings in general and of the particular (2) culture, (3) class, (4) caste, and (5) sex to which the individual belongs, and of the particular (6) group and (7) situation in which he happens to be at the moment; also his (8) individuality (personality). Accordingly, the author devotes almost the entire book to an examination of these eight variables (the last two chapters deal with research theory and methods).

The author believes that there is a body of integrated psychological theory—field, learning, psychoanalytic, and so on—applicable to those variables. Actually, he uses mainly the first two and especially the drive reduction learning theory.

Perhaps the reviewer lacks the anthropological and sociological background to evaluate critically the author's examination of culture, class, and caste, but to him that examination seems comprehensive, meaningful, and valuable. He is less enthusiastic, however, about the analysis of groups, sex character, personality, and especially of "the human-wide variable." Queener painstakingly follows a few authors and his own thinking about these topics through many twists and turns; but in that detailed analysis he sometimes seems to overlook equally fruitful approaches. In a discussion of leadership, for instance, he follows mainly the excellent lead of Lewin and his followers, but fails to use other equally fruitful recent work on leadership. This failure leads him to conclude tentatively that there is "a common element running through all types of leadership" (p. 275).

The "human-wide variable" is tentatively identified with instinct (p. 34). Four possible human instincts—self-preservation, maternal, competitive, and heterosexual—are analyzed and rejected as human instincts. Like many another social psychologist and sociologist, the author concludes that since there are no human instincts there is no

socially relevant behavior which you could predict if you knew only that an organism is human (p. 84). He does not say so, but he implies that this knowledge would not even help you in making such predictions. He does admit that innate determinants of human behavior are not entirely lacking (p. 84), but he does not elaborate upon that statement other than to indicate (pp. 74-78) that it may be inherently easier for human beings to learn certain behavior patterns than their opposites, viz., competition than cooperation.

There is little or no reference to human-wide variables which are the product of inevitable human learning in all environments, irrespective of culture (Dennis' autogenous learning). The author seems to be horrified, in fact, that anything might be inevitable about human nature. He makes no mention either of the universals found in all cultures, like speaking some language. In part, these omissions are deliberate; but, in part too, they seem to stem from overlooking much material of value in predicting social behavior.

This text, like most of those in social psychology, does not build on physiology or on general and child psychology in any comprehensive and thorough fashion. That was excusable in McDougall's day when general psychology was dominated by a sterile sort of introspectionism and there was no child psychology. But for the social psychologists of today to follow blindly the precedent set by McDougall and to develop their own inadequate versions of general and developmental psychology seems inexcusable. Social psychology texts need either to expand into two volumes, or, as seems wiser to your reviewer, to limit their domain.

These strictures apply, of course, to most texts in this area and not in any greater degree to this particular one than to the others. This is a systematic, well written, and richly illustrated text.

CLARENCE LEUBA.

*Antioch College.*

MORENO, J. L. *Sociometry, experimental method and the science of society: An approach to a new political orientation.* Beacon, N. Y.: Beacon House, 1951. Pp. xiv+220. \$6.00.

Moreno's reputation among American social scientists rests primarily on two specific contributions—the systematic measurement of patterns of choice and rejection among particular groups of people, usually referred to as sociometry; and the diagnostic and therapeutic use of role playing, usually referred to as psychodrama or sociodrama. Moreno himself, however, sees these contributions as secondary rather than primary; his major achievement, he feels, is the formulation of a new science of human relations, and the term "sociometry" in the present volume is used mainly in this expanded sense.

The book is made up of 34 papers, ten of which have not previously

been published. They range in time from 1923 to 1951 and in scope from very brief notes to long theoretical discussions. The best account of Moreno's general theoretical position is the concluding article, entitled "Origins and foundations of interpersonal theory, sociometry and microsociology." In this he states: "Sociometry aspires to be science within its own right. It is the indispensable prologue and preparatory science for all the social sciences. . . . It is not merely a slogan indicating a special type of research, a single method or a number of techniques."

A large portion of the book is devoted to explanation of the ways in which sociometry is different from and better than field theory, Gestalt theory, psychoanalysis, psychometrics, Marxism, and traditional experimental method. A few papers describe specific empirical applications of the sociometric approach. One excellent discussion, "Psychodrama and group psychotherapy: I," presents the essential features of psychodrama in capsule form.

Several articles written within the last few years justify the book's subtitle, "An approach to a new political orientation." Here we learn that "Sociometry is the sociology of the people, by the people and for the people," and also that "The revolutions of the socialistic-marxistic type are outmoded; they failed to meet with the sociodynamics of the world situation. The next social revolution will be of the 'sociometric' type." After these brave words it is a little disappointing to find that the most recent essay, dated 1951, advocates an essentially bureaucratic approach to sociometric reconstruction through the establishment of a department of human relations within the United States federal government.

It is probably evident by now that this reviewer is unable to take seriously either the claim that sociometry is a basic science coordinate with sociology, anthropology, and psychology, or that it offers any great promise for the worldwide alleviation of social ills. For this reason the present volume seems of considerably less value than other sociometric writings reporting the results of limited therapeutic efforts and specific empirical research.

JOHN HARDING.

*Commission on Community Interrelations of the American Jewish Congress.*

VOLKART, EDMUND H. (Ed.) *Social behavior and personality. Contributions of W. I. Thomas to theory and social research.* New York, Social Science Research Council, 1951. Pp. ix+338. \$3.00.

This volume consists of seventeen selections from the writings of W. I. Thomas. In addition, there is a thirty-two page introduction by the editor, a full bibliography of Thomas' papers and books, and a short biographical note that tells nothing of the man as a personality.

Since the present reviewer knew Thomas for thirty years, beginning as his graduate assistant in 1917, he prefers to comment on the editor's interpretations rather than to attempt more than brief notations regarding the content of the particular papers. If some of these comments seem a trifle severe, it is because this reviewer quite as much as the editor wants to see Thomas placed in proper perspective in sociology and social psychology.

Before making these comments, a word about the organization of the selections. The volume is divided into four parts.

1. The first, "Social science and social behavior," contains, among others on method, the much-discussed "Methodological Note" from *The Polish Peasant in Europe and America*, a joint work with Florian Znaniecki. It has never been made clear—either by Volkart or others—just how much of this "Note" was Thomas' own thinking and how much of it represented the interests of the co-author. Certainly the dichotomy "value" and "attitude" does not appear in other of Thomas' works and, as will be noted below, he had no interest in formal system building. Moreover, the "Note" was distinctly an afterthought and had no place in the design of the research in question.

2. The second part, "Social behavior and personal dynamics," contains, along with other papers, the classic statement on his theory of the four wishes and a suggestive piece from the *Polish Peasant*, "A theory of social personality."

3. The third part, "Social behavior and cultural dynamics," has several selections dealing with various aspects of social-cultural change. One of these, on the assimilation of immigrants, is from *Old World Traits Transplanted*, a book for which Thomas "was primarily responsible" but which was not made clear at the time it was published. (Two of his collaborators, Park and Miller, appear as the authors.) This is a recognition long overdue.

4. The final part contains an extended document "Outline for a program for the study of personality and culture." This was prepared for the Social Science Research Council and represents perhaps the first full review and appraisal of the field of culture and personality, just then emerging.

The most serious deficiency in this book is the almost complete neglect of Thomas' work prior to 1917, with the exception of one selection from his *Source Book for Social Origins* (1909). For the record it is well to indicate that the *Source Book* was one of the most important contributions of its time to sociology and social psychology. In this book, as in some of his earlier papers, Thomas made telling criticisms of the doctrine of recapitulation, of social evolutionism, and of the thesis of sharp biological differences among the races, all dogmas widely accepted at that time. Moreover, his criticism of what he aptly termed "particularism" was most appropriate. It gave clear evidence of his appreciation of multiple causation in relation to personality, society, and culture.

Another point is the editor's implication regarding Thomas' contribution to systematic theory. The editor confuses systematic theory with the making and using of hypotheses for research. As Volkart



well points out, Thomas was empirically oriented and his suggestions as to hypotheses and techniques for research were excellent for a period when most social psychologists and sociologists were still given to armchair theorizing. While deeply concerned with empirical studies, Thomas never, to my knowledge, showed any interest in system building. In fact, he always seemed to be temperamentally opposed to the construction of formal systems of thought.

A longer view of Thomas' development will show that in his early years he had a strong biological orientation, as is clear from certain of the early papers reprinted in *Sex and Society* (1907). Later he came up with his theory of the four wishes which beyond the merest statement was never systematically elaborated. All he did was to provide many case histories to illustrate each of the wishes in operation. Certainly the doctrine of the four wishes showed an appreciation of the subjective, inner elements in motivation, but they were not tied together logically with other theoretical considerations of human behavior. Later Thomas took up what he called "the situational approach." Yet, as Volkart points out, he never did resolve the implicit conflict between his earlier and later views regarding the interplay of subjective and objective elements in personality. But apparently this did not bother Thomas, who was not interested in logic-tight formulations. His chief concern was to expose the rich variety of human behavior.

Thomas' great confidence in the use of personal documents as basic source data on personality is clearly set forth in various of the reprinted selections. In this connection a further point may be made. At heart Thomas was an artist. He began his academic career in the field of literature, and it was from this background that he developed his concern with the personal document. While he saw the personal document as a source for certain hypotheses regarding important general aspects of behavior, he always had a strong curiosity about the idiosyncratic and even bizarre in human personality. In this regard he reminds one of those psychologists who stress the unique and autonomous features of personality rather than the more universal.

In closing this review I want to say that the editor has done an excellent job in indicating in his introduction and notes the many implications of Thomas' work for contemporary students of personality and culture. It has always concerned me that social psychologists whose roots lie in academic psychology have shown so little interest in, or knowledge of, the important work of W. I. Thomas. Let us hope that the publication of this collection of papers will at least help them to appreciate his place in the history of this important field.

KIMBALL YOUNG.

*Northwestern University.*



BOHM, EWALD. *Lehrbuch der Rorschach-Psychodiagnostik für Psychologen, Ärzte und Pädagogen*. Bern: Hans Huber, 1951. Pp. xxiii+406. Swiss fr. 36.

This textbook of Rorschach's perceptanalysis for psychologists, physicians and educators was written in German by a Dane and published in Switzerland. It aims to survey the present status of perceptanalysis without making the reading of Rorschach's Psychodiagnostics superfluous. Of necessity it repeats Rorschach's main ideas to make the new developments comprehensible. The book is based mainly on Swiss literature but it also uses Scandinavian and several American sources. The author tried to cover all the literature that was accessible to him. The survey is conscientious, the presentation is clear and succinct. The book offers an excellent opportunity to acquaint oneself with most European contributions including Bohm's own. Multicolored loose tables, in a separate appendix, aid in rapidly identifying the normal and small detail areas. However, there are no statistical frequency tables (Rorschach's own were not copied) but frequencies are indicated often in approximate numbers. It is better to have approximate figures that are valid than precise numbers that might not be valid. The thick book contains numerous details but they are well organized and cluster around the leading concepts.

The discussion of the administration of the test is one of the best in the literature. Bohm sounds like an expert teacher who is aware of the errors that students are likely to commit. His review of the Rorschach as an aid in neuropsychiatric diagnosis probably is the most complete available within the covers of one volume. The author relates the Rorschach well to general psychopathology and to clinical psychiatry, and thus, so to speak, presents the Rorschach in its natural habitat. The reader is continually warned of the tentative nature of many statements. It may be that Bohm would not have been reticent about some of the problems which he presents as unsolved if he had been more familiar with the pertinent literature in this country. He would not then have limited the function of the test to that assigned to it by its creator. Applying to the human personality terms used in explaining the functions of the nervous system, Bohm states that the Rorschach method provides a description of a "centrally conditioned (located in the brain) selection of impressions." In other words, the Rorschach enables us to make an inventory of socio-psychological stimuli which the individual perceives, consciously or unconsciously. According to Bohm, as well as Rorschach, the test can not—except under rare and favorable circumstances—tell us about the efferent behavior, about the overt reactions to the perceived stimuli, about the inhibitions which might interfere with overt reactions, or about the amount of available energy.

The author's conservative attitude, inimical to the spirit of reformation, is strong and deliberate. He stands by Rorschach and rejects consistently all major attempts to modify Rorschach's ideas. He does not even try to justify his rejection beyond saying that those modifications confuse. On the other hand, he is receptive to all innovations which leave all of Rorschach's own ideas intact. He corrects only when there is overwhelming factual evidence to support the correction, e.g., he rectifies Rorschach's notion that psychotics never produce a color shock. The book is provided with two indexes, a name index, and a particularly good subject-matter index. There is also a glossary of symbols in German, American, French, and Latin. Records of 31 subjects are reproduced with the scoring, the tabulation, a succinct diagnostic discussion, a brief personality analysis and a clinical summary. In the eight-page chapter on the theoretical foundations of the Rorschach test, Bohm lists experiments and observations which would be helpful in building a theory, although they were not made with the Rorschach method in mind. He cites studies on the emotional significance of chromatic colors, investigations of free association, figure and ground experiments, etc.

This readable book is rich in content. It has the merit of not misrepresenting the views of other investigators, even when the author disagrees with them. It is a good textbook in a field fraught with uncertainties and complex in its structure. It is a good aid to self-aid.

ZYGMUNT A. PIOTROWSKI.

*Columbia University and New York University.*

#### NOTICE

Dr. Wayne Dennis will become editor of the *Psychological Bulletin* as of January 1, 1953. All new manuscripts should *henceforth* be sent to him at the following address:

##### PSYCHOLOGICAL BULLETIN

Dr. Wayne Dennis, Editor

Department of Psychology

Brooklyn College

Brooklyn, New York

The Book Review Editor will be Dr. Edward Girden. All books and other publications intended for listing or review should be addressed to him as follows:

##### PSYCHOLOGICAL BULLETIN

Dr. Edward Girden, Book Review Editor

Department of Psychology

Brooklyn College

Brooklyn, New York

## BOOKS AND MATERIALS RECEIVED

ABRAMSON, HAROLD A. (Ed.) *Problems of consciousness*. (Transactions of the second conference, March 19-20, 1951.) Packanack Lake, N. J.: Josiah Macy, Jr., Foundation Publ., 1951. Pp. 178. \$3.25.

ADKINS, DOROTHY C., & LYERLY, S. B. *Factor analysis of reasoning tests*. Chapel Hill, N. C.: Univer. North Carolina Press, 1952. Pp. iii + 122. \$2.00.

ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of values*. (Rev. ed.) Boston: Houghton Mifflin, 1951. Pp. 12 (text) and 15 (Manual of Directions). \$.30.

BERDIE, RALPH F. *Concepts and programs of counseling*. Minneapolis: Univer. Minnesota Press, 1951. Pp. 81. \$1.75.

BOARD OF EDUCATION OF THE CITY OF NEW YORK. *A summary of the evaluation of the Aetna Roadometer performance test*. Divisional Bulletin No. 2. New York: Bureau of Educational Research, Division of Tests and Measurements, August 1951. Pp. 11.

BÜHLER, CHARLOTTE, LEFEVER, D. W., KALLSTEDT, FRANCES E., & PEAK, H. M. *Development of the basic Rorschach score supplementary monograph*. Los Angeles: Rorschach Standardization Study, 1952. Pp. iv + 71. \$2.50.

BÜHLER, CHARLOTTE, SMITTER, FAITH, & RICHARDSON, SYBIL. *Childhood problems and the teacher*. New York: Henry Holt, 1952. Pp. xi + 372. \$3.75.

CASSIRER, ERNST. *The philosophy of the enlightenment*. Princeton: Princeton Univer. Press, 1951. Pp. xiii + 366. \$6.00.

CLEUGH, M. F. *Psychology in the service of the school*. New York: Philosophical Library, 1951. Pp. vii + 183. \$3.75.

EKMAN, GOSTA, HUESEN, T., JOHANSSON, G., & SANDSTROM, C. I. (Editorial Committee). *Essays in psychology dedicated to David Katz*. Stockholm: Almqvist & Wiksell Boktryckeri AB, 1951. Pp. v + 283.

FENLASON, ANNE F. *Essentials in interviewing: For the interviewer offering professional services*. New York: Harper, 1952. Pp. xi + 352. \$4.00.

FESTINGER, LEON, & KELLEY, HAROLD H. *Changing attitudes through social contact*. Ann Arbor: Institute for Social Research, Univer. of Michigan, 1951. Pp. 83.

FISHER, BURTON R., & WITHEY, STEPHEN B. *Big business as the people see it: A study of a socio-economic institution*. Ann Arbor, Mich.: Institute for Social Research, Univer. of Michigan, 1951. Pp. iii + 200. \$1.50.

FLESCH, RUDOLF. *The art of clear thinking*. New York: Harper, 1951. Pp. vii + 212. \$2.75.

FLUGEL, J. C. *A hundred years of psychology*. New York: Macmillan, 1951. Pp. 424. \$3.25.

FORRESTER, GERTRUDE. *Methods of vocational guidance*. (Revised and enlarged.) New York: D. C. Heath, 1951. Pp. x+463. \$4.25.

FRENCH, THOMAS M. *The integration of behavior*. Vol. 1, *Basic postulates*. Chicago: Univer. of Chicago Press, 1952. Pp. xi+272. \$5.00.

FREUD, SIGMUND. *Psychopathology of everyday life*. New York: New American Library of World Literature, 1951. Pp. 168. \$.35.

GURVITZ, MILTON S. *The dynamics of psychological testing*. New York: Grune & Stratton, 1951. Pp. xii+368. \$.675.

HAVIGHURST, R. J., EATON, W. H., BAUGHMAN, J. W., & BURGESS, E. W. *The American veteran back home*. New York: Longmans, Green, 1951. Pp. vii+271. \$3.50.

HINCKELY, R. G., & HERMANN, LYDIA. *Group treatment in psychotherapy: A report of experience*. Minneapolis: Univer. of Minnesota Press, 1951. Pp. x+136. \$3.00.

HORNER, A. MUSGRAVE. *Speech training: A handbook for students*. New York: Philosophical Library, 1951. Pp. 176. \$3.75.

KEMPTHORNE, OSCAR. *The design and analysis of experiments*. New York: Wiley, 1952. Pp. xix+631. \$8.50.

HULL, CLARK L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951. Pp. viii+145. \$2.75.

HUMPHREY, GEORGE. *Thinking: An introduction to its experimental psychology*. New York: Wiley, 1951. Pp. vii+331. \$4.50.

HUSEN, TORSTEN, & HENRICSON, SVEN-ERIC. *Some principles of construction of group intelligence tests for adults: A report on construction and standardization of the Swedish Induction Test—the I-Test*. Stockholm: Almqvist & Wiksell, 1951. Pp. 100. 5 Sw. Cr.

JACKSON, EDITH B., & TRAINHAM, GENEVIEVE. *Family centered maternity and infant care*. (Supplement 1 to Problems of Infancy and Childhood, Transactions, Fourth Conference, 1950. Sponsored by Josiah Macy, Jr. Foundation, N. Y.) Caldwell, N. J.: Progress Associates, 1950. Pp. 29. 1-25 copies, 25 cents each; 26-100 copies, 15 cents each; 101 copies or more, 10 cents each.

JACOBI, JOLANDE. *The psychology of C. G. Jung*. New Haven: Yale Univer. Press, 1951. Pp. xx+244. \$3.75.

JEFFRESS, LLOYD A. (Ed.) *Cerebral mechanisms in behavior*, The Hixon symposium. New York: Wiley, 1951. Pp. xiv+311. \$6.50.

JOSEPH, ALICE, & MURRAY, VERONICA F. *Chamorroes and Carolinians of Saipan*. Cambridge: Harvard Univer. Press, 1951. Pp. xi+381. \$5.00.

KAPLAN, LOUIS, & BARON, DENIS. *Mental hygiene and life*. New York: Harper, 1952. Pp. xiv+422. \$3.50.

KELLY, E. LOWELL, & FISKE, DONALD W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951. Pp. xv+311. \$5.00.

KUHLEN, RAYMOND G. *The psychology of adolescent development*. New York: Harper, 1951. Pp. xvii+675. \$5.00.

LOEWY, HERTA. *The retarded child: A guide for parents, teachers and social workers*. New York: Philosophical Library, 1951. Pp. 160. \$3.75.

LUNDIN, WILLIAM H. *Projective techniques and psychotherapy: Examination of a therapy process through the use of the projective movement sequence and Rorschach technique*. New York: Society for Projective Techniques and Rorschach, 1951. Pp. 39. \$1.50.

MACDOUGALL, CURTIS D. *Understanding public opinion: A guide for newspapermen and newspaper readers*. New York: Macmillan, 1952. Pp. xii+698. \$5.00.

MCGEOCH, JOHN A., & IRION, ARTHUR L. *The psychology of human learning*. (2nd Ed., revised by Arthur L. Irion.) New York: Longmans, Green, 1952. Pp. v+596. \$5.00.

McKEAN, ELLY. *It's mine*. New York: Vanguard Press, 1951. (Pages unnumbered.) \$2.00.

MOLONEY, JAMES. *The battle for mental health*. New York: Philosophical Library, 1952. Pp. x+105. \$3.50.

PARSONS, TALCOTT, & SHILS, EDWARD A. (Eds.) *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1951. Pp. xi+506. \$7.50.

PIÉRON, HENRI. *Vocabulaire de la psychologie*. Paris: Presses Universitaires de France. Pp. 356.

PLOSCOWE, MORRIS. *Sex and the law*. New York: Prentice-Hall, 1951. Pp. ix+310. \$3.95.

POOL, ITHIEL DE SOLA. *The "Prestige Papers": A survey of their editorials*. Stanford, Calif.: Hoover Institute Studies, Stanford Univer. Press, 1952. Pp. 146. \$1.75.

POWELL, NORMAN J. *Anatomy of public opinion*. New York: Prentice-Hall, 1951. Pp. xi+619. \$5.50.

RAY, WILBERT S. *A laboratory manual for social psychology*. American Book Co., 1951. Pp. vii+173. \$3.00.

RALLI, ELAINE P. (Ed.) *Adrenal cortex*. (Transactions of the second conference, Nov. 16-17, 1950, New York, N. Y. Sponsored by Josiah Macy, Jr. Foundation.) Caldwell, N. J.: Progress Associates, 1951. Pp. 209. \$3.00.

RUNES, DAGOBERT D. *Spinoza dictionary*. New York: Philosophical Library, 1951. Pp. xiv+309. \$5.00.

SANDSTROM, CARL IVAR. *Orientation in the present space*. Stockholm: Almqvist & Wiksell, 1951. Pp. 193. 18 Sw. Cr.

SARASON, SEYMOUR B. *Psychological problems in mental deficiency*. New York: Harper, 1949. Pp. x+366. \$5.00.

SCHNEIDERS, ALEXANDER A. *The psychology of adolescence: A factual*



and interpretive study of the conduct and personality of youth. Milwaukee, Wis.: Bruce Publishing Co., 1951. Pp. xii+550. \$4.00.

SENN, MILTON J. E. (Ed.) *Symposium on the healthy personality*. (Transactions of Special Meetings of Conference on Infancy and Childhood, June 8-9 and July 3-4, 1950, New York, N. Y. Sponsored by Josiah Macy, Jr. Foundation.) Caldwell, N. J.: Progress Associates, 1950. Pp. 298. \$2.50.

SENN, MILTON J. E. (Ed.) *Problems of infancy and childhood*. (Transactions of the fourth conference, March 6-7, 1950. Sponsored by Josiah Macy, Jr. Foundation.) Caldwell, N. J.: Progress Associates, 1951. Pp. 181. \$2.25.

SHUTTLEWORTH, FRANK K. *The adolescent period: A graphic atlas and a pictorial atlas*. Monogr. Soc. Res. Child Developm. 1949, 14, Ser. No. 49, 1 & 2. Evanston: Child Development Publ., 1951. (Pages unnumbered.)

SLOTKIN, J. S. *Personality development*. New York: Harper, 1952. Pp. x+401. \$4.50.

SMITH, DEAN A., & WOODRUFF, M. F. A. *Deficiency diseases in Japanese prison camps*. London: His Majesty's Stationery Office, 1951. Pp. viii+209. SD 6/6.

SMITH, F. V. *The explanation of human behaviour*. New York: Macmillan, 1952. Pp. ix+276. \$2.75.

STETSON, R. H. *Motor phonetics: A study of speech movements in action*. Amsterdam: North-Holland Publishing Co. (for Oberlin College), 1951. Pp. xi+212. \$3.50.

STONE, C. P. (Ed.) *Annual review of psychology*. Vol. 3, 1952. Stanford, Calif.: Annual Reviews, Inc., 1952. Pp. ix+462. \$6.00.

TABA, HILDA, BRADY, ELIZABETH H., ROBINSON, JOHN T., & VICKERY, W. E. *Diagnosing human relations needs*. Washington: American Council on Educ., 1951. Pp. xi+155. \$1.75.

TINBERGEN, N. *The study of instinct*. New York: Oxford University Press, 1951. Pp. xii+228. \$7.00.

UNIVERSITY OF TORONTO (Staff of the Institute of Child Study). *Twenty-five years of child study*. Toronto, Canada: University of Toronto Press, 1951. Pp. xiii+182. \$4.00.

WATSON, ROBERT I. *The clinical method in psychology*. New York: Harper, 1951. Pp. xii+779. \$5.00.

WILLIAMS, GARDNER. *Humanistic ethics*. New York: Philosophical Library, 1951. Pp. xii+ \$3.75.

YOUNG, Pauline V. *Social treatment in probation and delinquency. Treatise and casebook for court workers, probation officers and other child welfare workers*. (2nd ed.) New York: McGraw-Hill, 1952. Pp. xxvi+536. \$7.00.

*L'année psychologique*. (Cinquantième année. Vol. jubilaire offert en hommage à Henri Piéron.) Paris: Presses Univers. de France, 1951. Pp. xvi+718.

ukee,

ality.

Child-

ed by

ciates,

dhood.

nsored

ciates,

c atlas

9, 14,

Pages

1952.

ses in

Office.

York.

ents in

n Col-

1952.

0.

Vick-

Amer.

Univer.

study).

oronto

York

pphical

quency

er child

xxvi +

e offer

, 1951